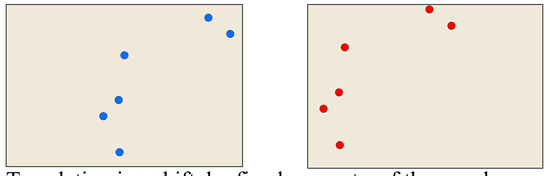
2

GEOMASKING

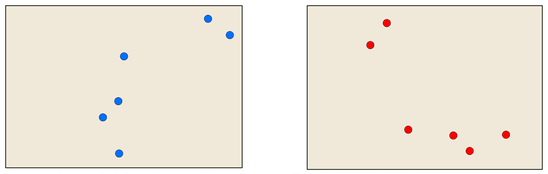
Geomasking refers to techniques that preserve the confidentiality of individual health records while still permitting some spatial analysis. Examples include data aggregation, surface generation, transformation, which is a non-random point shift, and random perturbation, a random point shift. The end user must be informed when data have been masked.

Transformation Examples

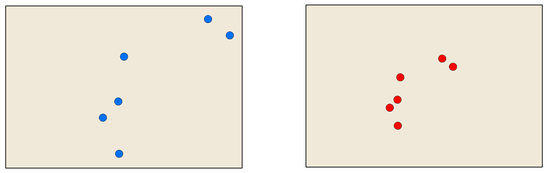
ORIGINAL GEOMASKED



Translation is a shift, by fixed amounts, of the x and y coordinates.



Rotation shifts the points by a fixed angle about a pivot point.

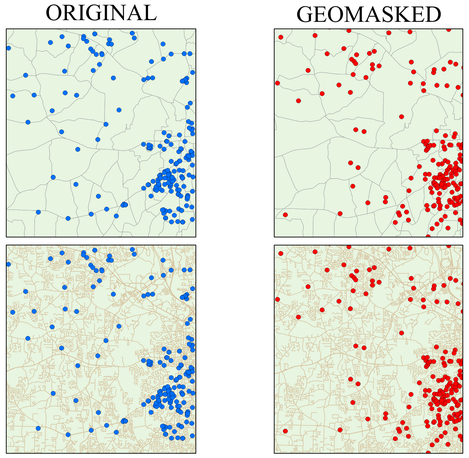


Change of scale transforms the distances, by a constant, between each point.

Random Perturbation Examples

A quick method is to shift each point randomly in the block group within which it is located. Census BGs contain from 600 to 3,000 people with an optimum size of 1,500. BGs vary in area; less densely populated BGs are larger in area than more densely populated BGs. Therefore, points in less populated areas have the potential for a greater shift.

ORIGINAL GEOMASKED



1

Confidentiality and Geocoded Health Data

PURPOSE

Geocoding is of great value for mapping and analysis; however, issues of confidentiality arise when working with health data.

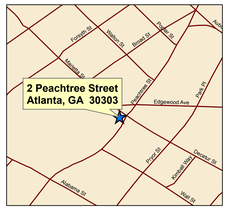
Methods to comply with the HIPAA Privacy Rule, for both presentation mapping and digital data distribution, are explored.

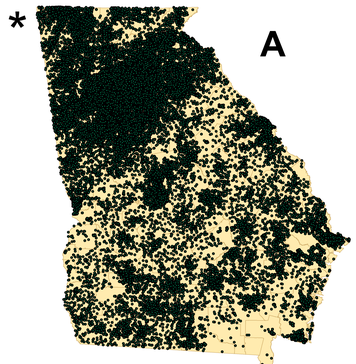
WHAT IS GEOCODING?

Address geocoding, a commonly performed GIS operation, involves matching street addresses to geographic coordinates, usually longitude and latitude, on a map.

Many options exist for mapping the geocoded data for presentation. For example, as in the figures of 2004 births below, maps can display:

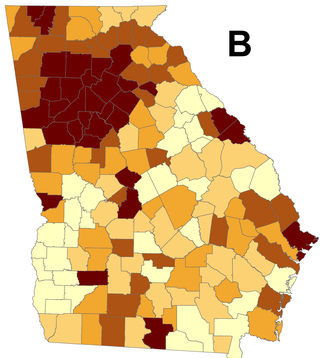
The resulting point locations are useful for many different kinds of spatial analysis, including health-related applications.



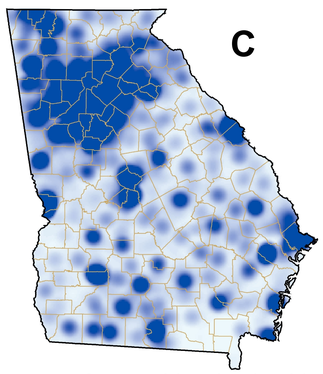


Individual points (e.g. displaying mother's residence)

*** All example data are either de-identified or are *not* Protected Health Information.**



Enumeration units, such as census tracts, counties, or public health districts, with aggregated point totals or calculated rates, percentages, densities, etc. (e.g. births per square mile by county)



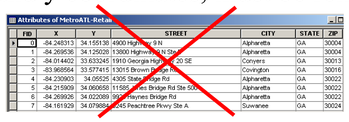
"Surfaces" with point densities calculated (e.g. births per square mile as calculated directly from points)

CONFIDENTIALITY ISSUES

Although health data mapping is not addressed directly in the HIPAA Privacy Rule, pinpointing a location can enable "ambitious de-identifiers" to determine an individual's identity, thereby violating the Privacy Rule. Confidentiality issues are different depending on what is digital spatial data, or a map, either electronic or hardcopy.

Digital Spatial Data

GIS data include attribute tables that show characteristics of map features. Identifying attributes, such as names, street addresses, or x and y coordinates, must be removed.



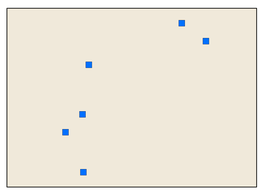
Reverse Geocoding is a concern. Given a point location, GIS software returns x and y coordinates which can be used to obtain an address.

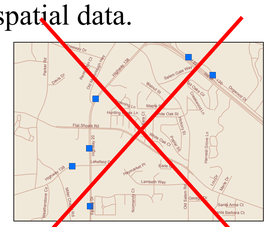
Mapping to Protect Privacy

Map points only at small scales. In Map A above, the scale is too small to permit identification of individuals.

Aggregate data to a geographic entity such as the county and then map rates as shown in Map B. Aggregation, a type of geomasking, also preserves privacy when distributing digital spatial data.

Create a density surface as in Map C. Surfaces, also a type of geomasking, preserve privacy when distributing digital spatial data.



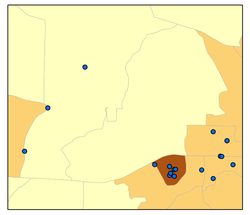


At large scales, map data without locational reference,such as streets, boundaries, or landmarks

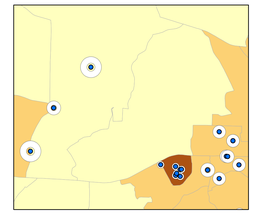
3

Random Perturbation (Continued)

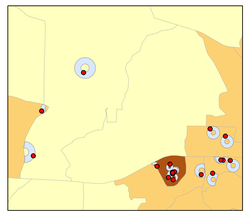
A better method is to shift each point randomly within a donut-shaped buffer sized to reflect the local population density. This method offers more control in balancing the masking of locations with preserving the spatial distribution. The average distance between people for the target population, = √1/Population Density, is calculated for each point. The results are multiplied by 1 or 2 and by 3, 4, or 5 to create the inner and outer donut ring distances, respectively.



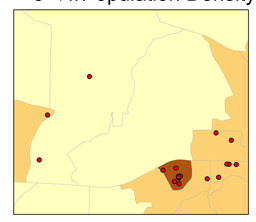
Original points overlaid on population density. Darker shades mean higher density.



Inner ring
= 2*√1/Population Density.
Outer ring
= 5*√1/Population Density.



Donuts clipped to block group boundaries. Point placed randomly within each donut.



Geomasked points

REFERENCES

Armstrong, M.P., Rushton, G., and Zimmerman, D.L. (1999). "Geographically masking health data to preserve confidentiality." Statistics in Medicine **18**: 497-525.

NIH. "Research Repositories, Databases, and the HIPAA Privacy Rule." Publication Number 04-5489. January 2004. <http://privacyruleandresearch.nih.gov/research_repositories.asp>.

Stinchcomb, Dave. "Procedures for Geomasking to Protect Patient Confidentiality." ESRI International Health GIS Conference. Washington, DC. October 2004. <http://gis2.esri.com/library/userconf/health04/papers/pap3012.pdf>

AUTHOR

Elaine J. Hallisey, MA
Georgia Department of Human Resources
Division of Public Health
Office of Health Information & Policy
2 Peachtree Street, Atlanta, GA 30303
Email: ejhallisey@dhr.state.ga.us

Publication Document Tracking Number: DPH07/081HW