Geovisualization of Attribute Uncertainty

Hyeongmo Koo[1], Yongwan Chun[2], Daniel A. Griffith[3]

University of Texas at Dallas, 800 W. Campbell Road, Richardson, Texas 75080,
[1]Email: hxk134230@utdallas.edu
[2]Email: ywchun@utdallas.edu
[3]Email: dagriffith@utdallas.edu

Abstract

Visualization of attribute uncertainty is often necessary for a successful data analysis. However, visualization tools for uncertainty are barely available in a GIS environment. This paper aims to propose a framework for visualization methods to represent attribute uncertainty and to prototype an implementation of these methods in a standard GIS environment. These tools that are implemented using ArcGIS Engine and C# are demonstrated with ACS data.

Keywords: GIS, uncertainty, data quality, visualization, bivariate mapping.

1. Introduction

Spatial data uncertainty or data quality potentially has an impact on data analysis, and hence, understanding uncertainty in data is often necessary for a successful data analysis. Because uncertainly is often embedded in the attributes of geographic features, a visual representation of attributes does not necessarily present an underlying pattern in a phenomenon, and often requires uncertainty of the attributes additionally. However, simultaneous representation of these two types of information (i.e., attributes and their uncertainty) is challenging. Many researchers have investigated visualization methods to effectively and efficiently represent spatial data uncertainty (e.g., MacEachren et al. 1998; MacEachren et al. 2005; Sun and Wong 2010). However, visualization of uncertainty is still not much popular in a geographic information system (GIS) environment with several exceptions such as Heuvelink et al. (2007) and Pebesma et al. (2007). This paper aims to propose a framework for visualization methods to represent attribute uncertainty and to prototype an implementation of these methods in a standard GIS environment.

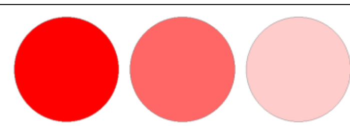
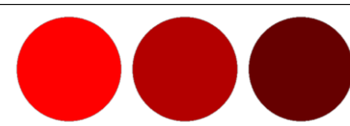
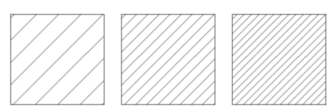
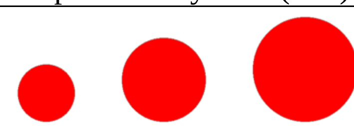
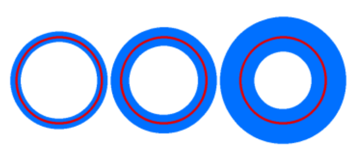
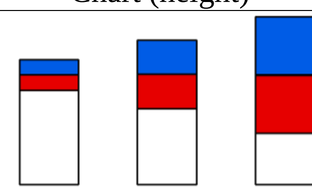
2. Method

Attribute uncertainty often is represented with another map in addition to an attribute map (e.g., Leitner and Buttenfield 2000), with this map juxtaposed to the attribute map for visualization purposes (MacEachren et al. 1998; Kardos et al. 2003). In a different perspective, an attribute and its uncertainty can be considered as two variables, allowing them to be subjected to bivariate mapping techniques. In bivariate mapping, overlay methods allow a single choropleth map to represent an attribute that is overlaid by uncertainty shown with different symbols on top (e.g., MacEachren et al. 1998; Sun and Wong 2010). A popularly used approach is to utilize coloring characteristics, where saturation, hue, and lightness are used to visualize attribute uncertainty (e.g., Hengl

2003). Also different symbol sizes are a widely used visual variable to represent attribute uncertainty.

Table 1 presents a framework for combinations of visual variables to simultaneously represent an attribute and uncertainty. The first approach is to use coloring properties method to represent uncertainty and proportional symbol for an attribute. In this configuration, a graduated or proportional symbol map is more appropriate to represent a count or frequency variable. In previous studies, saturation, crispness of symbol edge, or transparency depicts attribute uncertainty information, with graphical symbol sizes representing an attribute (MacEachren et al. 2005). Specifically, uncertainty is represented by saturation and value in Hue-Saturation-Value (HSV) color model, and the saturation or the value proportionally decreases from those of a base symbol as uncertainty increases.

Table 1. Framework of Uncertainty visualization methods

Attribute	Uncertainty (low ↔ high)		
	Categories	Visual variables	
Proportional symbol	Coloring	Saturation	Value
			
Choropleth map	Overlaid symbol	Texture (separation)	Proportional symbol (size)
			
Symbol size or height between confidence bounds	Integrated symbols	Proportional symbol (size)	Chart (height)
			

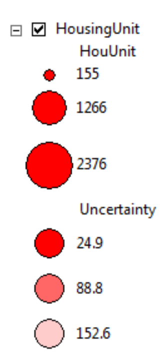
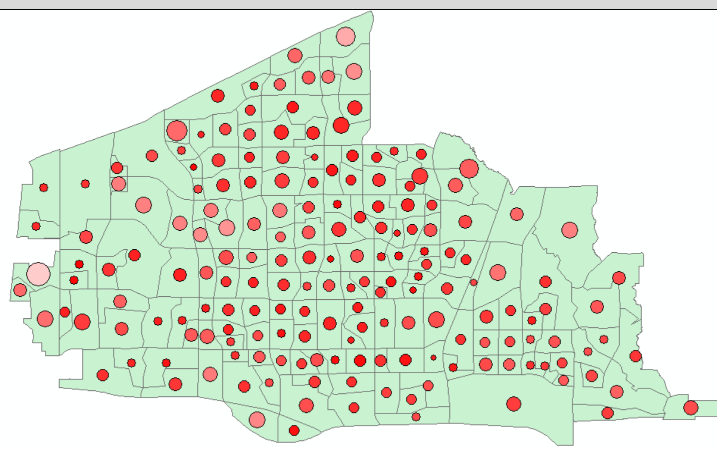
Second, overlaid symbols on a choropleth map are implemented to visualize attribute uncertainty. A choropleth map is suitable to represent an interval or ratio scale variable, and symbol color on a choropleth map represents an attribute. Overlaid symbols, such as textures and circles, are used to visualize attribute uncertainty. MacEachren et al. (1998) used a choropleth map overlaid with texture symbol approach to represent reliability of data. Sun and Wong (2010) use separations of line fill in polygon symbols, i.e., line texture, to represent degrees of attribute uncertainty for which an implementation is available in a form of ArcGIS Add-in tool. Also, proportional symbol sizes can be set with uncertainty level.

Third, uncertainty can be integrated in symbols of an attribute. In this approach, uncertainty is represented as interval around an attribute. A composite symbol can be utilized, which is barely available as built-in function in GIS packages. For example, around a circle symbol to represent an attribute, other symbols are integrated to represent 95% confidence interval as in Table 1. In chart symbol, a red box and a blue box respectively represent lower and upper confidence bound at given confidence level. The common boundary between a red box and a blue box represents an associate attribute. These methods allow users to visually inspect a possible range of an attribute with an associated uncertainty level.

3. Application

These uncertainty visualization methods are demonstrated using 5 year American Community Survey (ACS) data for the city of Plano in Texas. The ACS data provides a margin of error (MOE) as well as surveyed attributes, which is based upon the standard error of the sampling distribution for an attribute variable. Since MOEs represent the lower and upper bounds with 90-percent confidence level, the standard error of an attribute can be directly calculated from the MOE: that is, MOE = 1.645 * standard error (U.S. Census Bureau 2009). . Here, two variables with their MOEs are obtained from the 5-year ACS estimates (2009-2013) at the census block group level. Median age was chosen to illustrate the choropleth map with overlaid symbols, and the number of housing units is chosen to illustrate proportional symbol with coloring properties and overlaid symbols with different sizes. Visualization tools for attribute uncertainty are implemented with C# with .Net Framework 4 based on ArcEngine 10.1.

Figure 1 shows the results of proportional symbol with coloring properties. Graphical symbol sizes represent the number of housing unit, and associated uncertainties are depicted with saturation and value in the HSV color model. Low saturation and value represent a high level of uncertainty. When the symbol sizes are relative larger for the sizes of spatial units, these symbols can be overlapped with each other. Thus, the size of symbols may need to be chosen carefully; here, the drawing order of the symbol is set as from the largest symbol to the smallest symbol; that is, small symbols are drawn on top of large symbols. .

Visual variables	Legends	Results
Saturation		

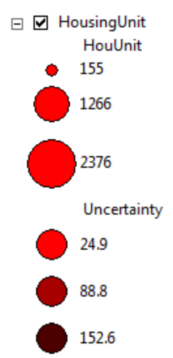
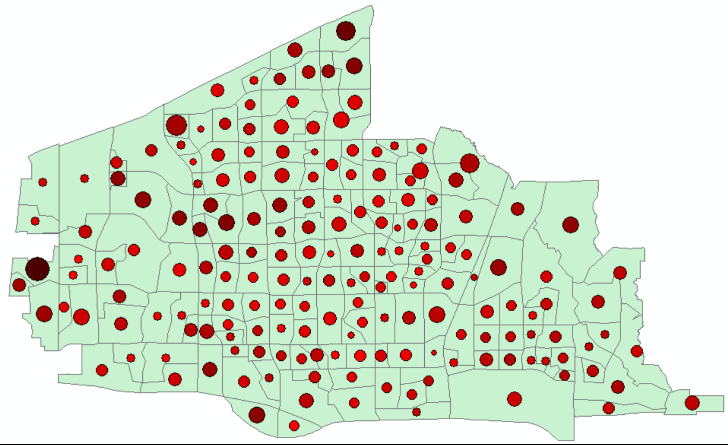
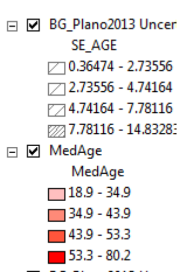
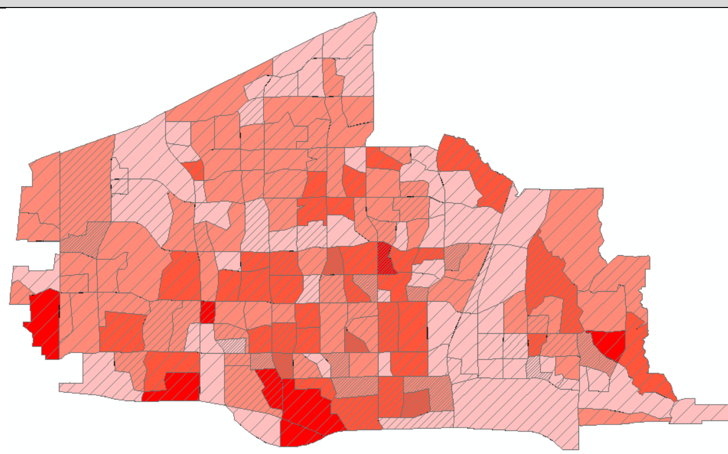
<table>
<tr><td>Value</td><td>
</td><td></td></tr>
</table>

Figure 1. Proportional symbols with coloring properties method for medium housing units with their standard errors

Figure 2 shows an uncertainty visualization using the choropleth map with overlaid symbols for median age with their standard errors. The colors of the choropleth map represent the attributes, and the separations of line fill symbol depict attribute uncertainty. Small line separations in the line fill symbol textures describe high levels of uncertainty. With proportional symbology, symbol sizes can furnish another approach to represent uncertainty levels; big symbols describe high level of uncertainty. These two approaches have a common challenge that it is difficult to recognize classes, when small areal units is described with a small number of line hatches; i.e., large separations in line fill symbol. Therefore, the separation in line symbols should be carefully chosen to effectively represent high and low uncertainty regions with a single map

<table>
<tr><th>Overlaid Symbol</th><th>Legends</th><th>Results</th></tr>
<tr><td>Textures</td><td>
</td><td></td></tr>
</table>

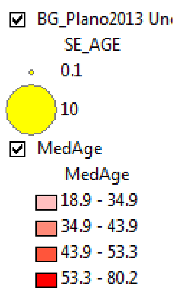
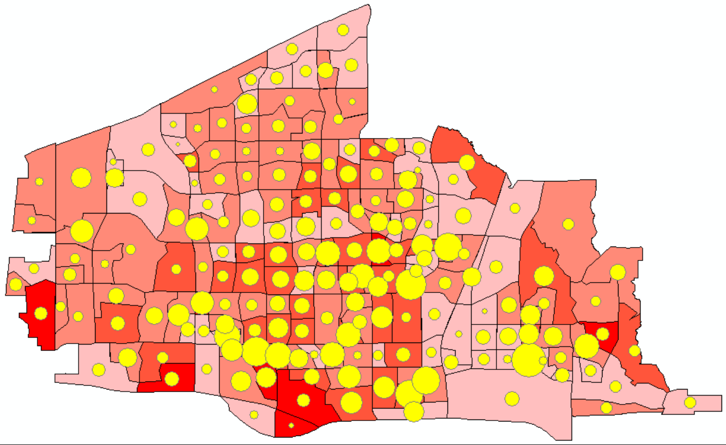
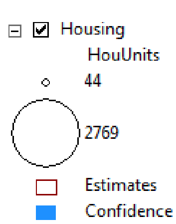
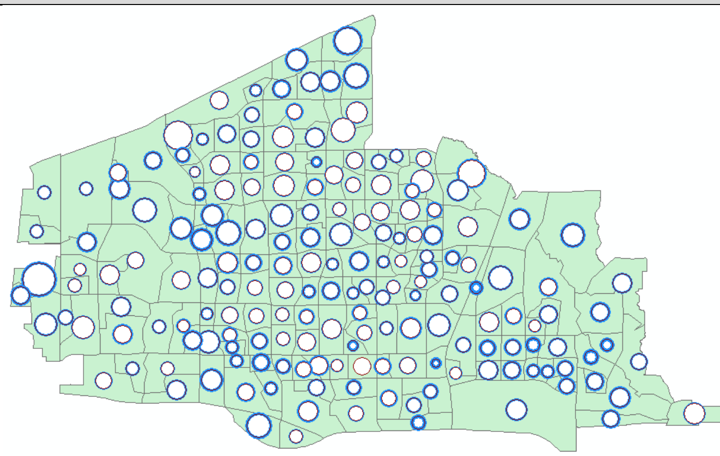
Proportional symbol		

Figure 2. A choropleth map with overlaid symbols and its results for median age estimates with standard errors

Figure 3 shows the number of housing units with the 95 percent confidence intervals, which are integrated in individual symbols. The upper map is prepared with overlaid symbols in different symbol sizes and the lower map show a chart map with a bar. In the upper map, the red lines in the proportional symbols represent the attributes, and the blue bounds describe their confidence interval. Thus, a thicker blue bound indicates a higher uncertainty level of the attributes. For the chart symbol map, the blue and red stacked chart represent respectively the upper and lower bounds of the attributes. The common boundary between the red and blue portions represents an attribute. Hence, the number of housing unit estimates is represented by the sum of the white and the red portions. The level of uncertainty is represented with the heights of red and blue portions.

Overlaid Symbol	Legends	Results
Proportional Symbol		

Chart

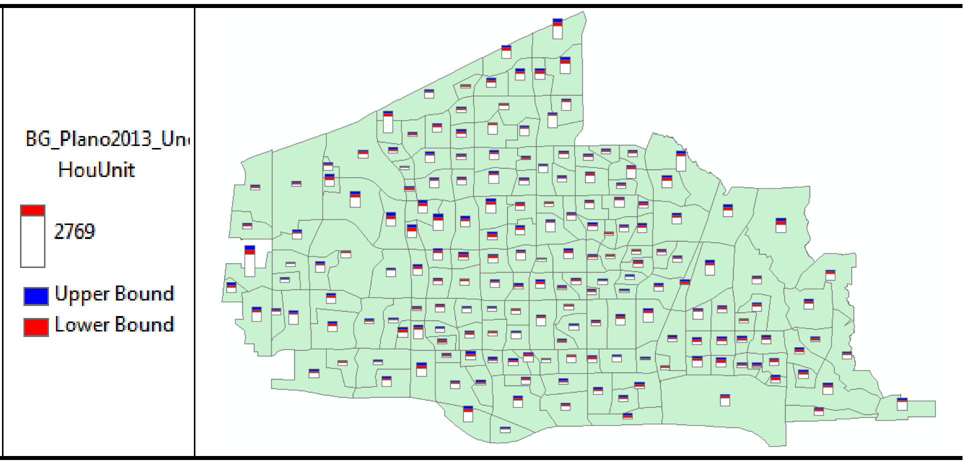


Figure 3. Overlaid symbols with different sizes and the resulting map for number of housing units, with 95 percent confidence intervals

4. Conclusion

Geovisualization of spatial data can be improved by including their attribute uncertainty that assists people to recognize an underlying pattern of data. However, there is still shortage in availability of such tools in a GIS environment, some existing tools in GIS packages utilize visualization methods based on bivariate mapping techniques. These methods can be categorized into three categories, which are proportional size of symbol with coloring properties, a choropleth map with overlaid symbols, and integrated symbols. Implementations of such tools in a GIS environment can enhance understanding of data by representing attributes and uncertainty.

These visualization methods often need a further adjustment. In a proportional symbol map, symbols can overlap with each other, and line texture symbols may not clearly represent attributes and their uncertainties for small areal units. Hence, minimum and/or maximum symbol sizes needs to be carefully set in a proportional symbol map. Similarly, the size of line separation in line fill symbol map needs to be carefully set. Also, symbol sizes in an integrated symbol depend on attributes and their confidence intervals. This method is more useful for variables with large standard errors. Future research requires evaluating the visualization methods.

5. Acknowledgements

This research was supported by the National Institutes of Health, grant 1R01HD076020-01A1; any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Institutes of Health.

6. References

Hengl, T. 2003. Visualisation of uncertainty using the HSI colour model : computations with colours. In *Proceedings of the 7th International Conference on GeoComputation*. Southampton, United Kingdom.

Heuvelink, G. B. M., J. D. Brown, and E. E. van Loon. 2007. A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science* 21 (5):497-513.

Jiang, B., F. Ormeling, and W. Kainz. 1995. Visualization support for fuzzy spatial analysis. In *Proceedings of ACSM/ASPRS Conference*, Charlotte, North Carolina, U.S.

Kardos, J., A. Moore, and G. Benwell. 2003. Visualising uncertainty in spatially-referenced attribute data using hierarchical spatial data structures. In *Proceedings of the 7th International Conference on GeoComputation*. Southampton, United Kingdom.

Leitner, M., and B. P. Buttenfield. 2000. Guidelines for the Display of Attribute Certainty. *Cartography and Geographic Information Science* 27 (1):3-14.

MacEachren, a. M., C. a. Brewer, and L. W. Pickle. 1998. *Visualizing georeferenced data: representing reliability of health statistics. Environment and Planning A* 30:1547-1561.

MacEachren, A. M., A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science* 32 (3):139-160.

Pebesma, E. J., K. de Jong, and D. Briggs. 2007. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. *International Journal of Geographical Information Science* 21 (5):515-527.

Sun, M., and D. W. S. Wong. 2010. Incorporating Data Quality Information in Mapping American Community Survey Data. *Cartography and Geographic Information Science* 37 (4):285–299.

U.S. Census Bureau. 2009. *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*. Washington, DC: U.S. Goverment Printing Office.