

# VISUALIZATION OF VAGUE CATEGORY COUNTS - INTRODUCING THE FUZZY DOT DENSITY MAP

Ola Ahlqvist  
154 N. Oval Mall  
Columbus, OH 43210, U.S.A.  
[ahlqvist.1@osu.edu](mailto:ahlqvist.1@osu.edu)

## **ABSTRACT**

This paper argues that mapping and visualization needs to consider different approaches to the various aspects of geographic information uncertainty. Of the three main types of uncertainty identified here – error, vagueness, and ambiguity – vagueness is an inherent property of many geographic phenomena and I argue that its mapping requires a specific visual metaphor. A survey of the uncertainty visualization literature reveals that very few have targeted visualization of vagueness, and the dot map is identified as a viable candidate for integral visual portrayal of quantity, pattern and vagueness. In a synthetic example contextualized within the application area of poverty mapping I elaborate on some general design options suggested in the literature; color saturation, transparency and object crispness. The examples demonstrate that the proposed fuzzy dot mapping technique can employ these visual variables and could therefore provide a viable technique for the representation of quantities of vague phenomena. The presentation also highlights some of the issues with the different approaches related to visibility and resolution acuity.

## **BACKGROUND**

Uncertainty is an inherent property of most geographic data (Golledge 2002) and much research has been directed towards its conceptualization, and description. Early efforts to conceptualize spatial data uncertainty resulted in a recognition of accuracy, resolution, consistency, and completeness (Verigin 1999). The past decade or so has witnessed increased research in the modeling of uncertainty and the way uncertainty propagates through data processing, analysis and interpretation. An important development is that the previously tight association between uncertainty and probability theory has been replaced by recognizing several facets of uncertainty. Many authors now follow basically the same major typology and distinguish *error*, *vagueness* and *ambiguity* as separate forms of uncertainty (Klir & Wierman 1999; Fisher 1999; Zhang & Goodchild 2002; Leyk et al. 2005). At the same time there has also been an increased interest in illustrating information uncertainty together with the mapped data.

Conceptually, adding uncertainty information to a map display is similar to adding another variable, but visualization of spatially distributed uncertainty poses a major problem since typically two or more visual dimensions already have been used to display the data (Bastin et al. 2002). Consequently most work on uncertainty mapping have called for some form of bi-variate or multi-variate mapping technique and examples

include the use of separate maps – one map for data and one for uncertainty (e.g. Burrough et al. 1992) , single maps with bi-/multivariate symbology (MacEachren et al. 1998) , animation (e.g. Fisher 1996) , and data exploration techniques where a number of tools (including the above) are combined in a single interface (Anselin 1999; Gahegan et al. 2002) . Despite this wealth of research there are still no clear guidelines for how to visually represent uncertainty (MacEachren et al. 2005) . The closest to a concrete guide for cartographic visualization of uncertainty can be found in the combined works of MacEachren (1995), Buttenfield and Weibel (1988) ; Buttenfield and Beard (1994) , but unfortunately they do not reflect the most recent developments of uncertainty conceptualization and representation summarized in the previous paragraph. For work on uncertainty visualization there is a need to recognize that uncertainty is a very general concept, that it can be separated into several facets, and that each facet may require unique symbolization approaches. From that perspective, and using the typology above we find that most work on geographic uncertainty visualization has been concerned with error. Less attention has been given to uncertainty about the actual existence of a phenomenon or the graded character of vague objects, related to ambiguity and vagueness respectively.

Fisher (2000) argues that many geographic objects, relations, and processes are well defined as to their central concept but surrounded by broad zones of transition and vagueness. He also proposes that vagueness can be identified by testing if a concept is susceptible to the Sorites Paradox (Ibid.). The formulation of the Sorites Paradox goes back to antiquity and is often exemplified by considering the concept of a heap. One grain of sand is obviously not a heap and adding one more grain does not make it a heap. If we repeatedly add one grain we will eventually form a heap of sand, but at what point did it become a heap? The same paradox can be found in the concepts of e.g. poverty, close, east, and downtown. We find that it is hard if not impossible to define an exact limit for what poverty is, what distance is close, etc. Standard set theory and logic fails to acknowledge the graded nature of these geographic notions, but fuzzy sets and fuzzy logic (Zadeh 1965) has the ability to address vagueness and degrees of truth in geographic information (V. B. Robinson 2003; Fisher 2000) . In fuzzy set theory we replace the boolean logic, either belonging or not belonging to a set, with a degree of belonging to a fuzzy set often represented by a real number from 0 to 1.

The remainder of this paper will address the mapping of counts or quantities of vague or Sorites susceptible objects. I will do so using a worked example of poverty mapping in which I initially demonstrate that commonly used mapping techniques rarely attempt or manage to represent the vagueness of poverty.

## **MAPPING OF VAGUE QUANTITIES – A POVERTY MAPPING EXAMPLE**



Figure 1 Examples of two standard mapping techniques, choropleth (left) and proportional symbol (right), to show the spatial pattern of poverty for some hypothetical Recent works have reported about the challenge of mapping poverty in a way that can truthfully support our understanding of the geographic pattern that poverty makes and to guide policy interventions (CIESIN 2006; Davis 2003). To produce maps, poverty is commonly translated into some form of monetary measure such as the headcount ratio or the poverty gap ratio (Hyman et al. 2005) often defined as the proportion of the population living in households whose per capita expenditures is below the “overall poverty line” according to some definition. These interpretations of poverty lend themselves well to standard choropleth or proportional symbol mapping techniques, and several examples of such poverty maps can be found in recent literature (CIESIN 2006; Glasmeier 2006; Hyman et al. 2005). For example, the U.S. Census Bureau measures poverty as a ratio of family income to a poverty threshold based on family size and ages of family members. (<http://www.census.gov/hhes/www/poverty/poverty.html>). Census data can then be aggregated and reported for enumeration units as a count or percentage of households below a poverty threshold. Figure 1 shows two mapping methods frequently used in the poverty mapping literature using an example of six hypothetical census enumeration areas.

One problem with these maps is that they enforce a binary interpretation of poverty based on a crisp cut-off level which is at odds with the graded nature in which poverty is conceptualized and measured (c.f. Davis 2003) . You can be very poor, or you can be poor but not extremely so, and then there are all the “non-poor” people above a defined threshold who still are in hardship and in grave risk of falling into poverty with just one big event, e.g. an unexpected bill or a hurricane hitting the area. Still, many poverty indicators such as infant mortality rates and household income mostly end up being mapped as crisp, non-contestable metrics, producing an unnecessarily simplified picture of this Sorites susceptible phenomenon. Maps such as the two above only represent quantities below a poverty threshold and do not show any information about the distribution of populations around this cut-off level.

To illustrate the problems this can produce in terms of using these maps for policy purposes Figure 2 illustrates some possible distributions of poverty metrics for households that would be expressed as the previous maps. In this figure we can see that completely different poverty situations can look identical in a regular map or very similar situations can look very different.

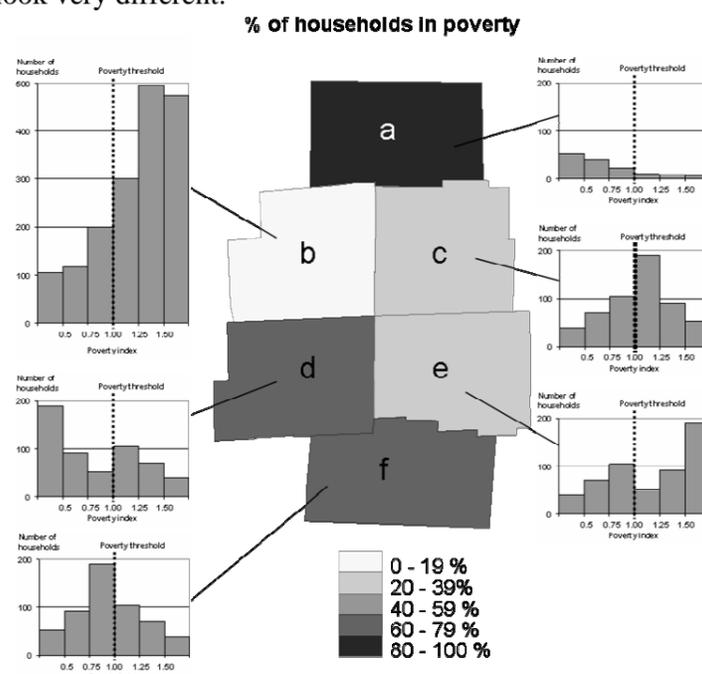


Figure 2 The map and histograms illustrates how different underlying statistical distributions will be generalized by a choropleth map using a crisp threshold for determining the quantity of a category, in this case people in poverty.

For example, the two lower left enumeration units (Figure 3 d and f) visually look the same but the underlying distribution shows a large number of households in severe poverty in the upper unit (d) whereas most of the households in the lower unit (f) is just below the poverty threshold. The map suggests a difference in poverty for enumeration units (c) and (f), but the distributions are very similar. The only reason for the big visual difference is how the location of the poverty threshold cuts just below or just above the mode of each distribution. Furthermore, the big difference between units (a) and (b) in the Figure 1 maps is because of the use of a ratio that does not reflect the absolute number of people in poverty. Evidently, none of these standard maps is able to visually capture the rich variation in the underlying data and the notion of poverty as a graded phenomenon.

An alternative approach would be the use of a fuzzy set membership function to allow for a graded belonging to the poor category. We could the for example set the federal poverty threshold to correspond with a membership value of 0.5 which could be translated as a

borderline case between poor and not poor and thus a partial member of the fuzzy set poor. Income-to-poverty ratios lower than that would have higher membership values and vice versa. Membership values of 1.0 would be considered full members of the poor fuzzy set and values of 0 would be outside the set. One of the key issues in using fuzzy sets for a geographic problem is the specification of grades of membership and a useful review of these issues together with some recommendations is provided by Robinson (2003) . However, there remains little guidance from the literature as to what mapping technique would lend itself best to the portrayal of vague category counts like these. Among the most common thematic mapping techniques; choropleth, isarithmic, proportional symbol, and dot density maps, only the latter two are generally recommended for absolute quantities or counts. In this paper I will focus attention on the dot density technique as it uses a particularly versatile symbology, capable of representing both absolute values as well as revealing the overall pattern of a distribution, and the technique is also easily extended to represent multiple attribute dimensions (Coulson 1990) . Other mapping techniques are possible and is currently under investigation. In the following section I will propose a modification of the common dot map to accommodate for uncertainty in the absolute count of vague geographic phenomena such as poor households.

### **THE DOT MAP**

Generally the Dot Map technique is employed to depict the density of a variable over an area. Eight different variations of the dot mapping technique were identified by Dahlberg (1967) , but the most common version in use today is where each point is placed on the map to represent a standard unit value, for example so that 1 dot equals 100 people. The technique has mostly been used for population density and agricultural production, but there are many examples of creative modifications and uses of the dot mapping technique to depict e.g traffic flow (Schultz 1961; Ullman 1949) and topographic relief (Pawling 1973) . Coulson (1990) listed several arguments in favor of the dot map over other standard mapping techniques (choropleth, isopleths, and graduated symbol maps) and mentioned its power to reveal overall patterns as the main benefit.

We previously saw that including uncertainty information in a quantitative map essentially adds an extra attribute dimension. Dot mapping has been recognized as a possible technique for multi-variate data and several studies have explored its use and effectiveness (e.g. Coulson 1990; Rogers & Groop 1981) . Still, the use of multi-variate dot density maps is rare and its use for portraying vagueness is almost non-existent. The only example known to the author is David Plane's (1998) dot maps that depict fuzzy set membership for entire enumeration units. Using state-level data on U.S. migration patterns he argued that dot density maps is preferred over choropleth and isarithmic maps because the dot patterns creates a better visual depiction of the graded character of fuzzy regions. Gradedness was represented by more or less dense dot patterns and was not related to an underlying quantity. As such his design is not able to represent an underlying quantity, let alone a quantitative distribution, for each enumeration unit.

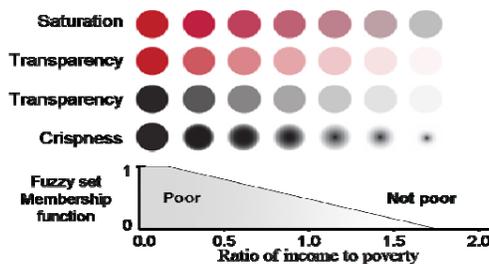


Figure 3 Four visual approaches; Saturation, Transparency (b/w and color), and Crispness, applied to dot symbols for representation of fuzzy set membership values.

### THE FUZZY DOT MAP

The purpose of the proposed fuzzy dot map is to visually represent a distribution of Sorites susceptible phenomena. Using the poverty example again, and assuming we have data on the number of people and an associated poverty metric, the fuzzy dot map will seek to use a well contrasting symbology for the number of people in severe poverty, less contrasting for counts of less poor, and so on, such that counts for people increasingly above the poverty limit should be represented by a symbology that almost blend into or make up the background. In this way the visual representation would supposedly correspond to the notion of poverty as a graded phenomenon. The visual portrayal most suitable for fuzzy membership values has not been established but some suggestions exist (Gershon 1998; Drecki 2002; Nelson 2000; MacEachren et al. 2005; Hengl 2003). In summary, these works suggest that visual crispness, saturation, or transparency would be potential candidates for representing vagueness but it is unclear how these apply to the dot map.

A fuzzy set membership function was established to guide the visual representation of dots that correspond to quantities of population with a partial membership in the fuzzy set poor. The U.S. Census Bureau tabulation (P88) reports for each census unit the total population for whom poverty status has been determined and separates the population into nine classes of poverty ratio ranges. Based on these we can then scale the dot colors, transparencies and crispness correspondingly. The visual symbols in Figure 3, applied to the maps in Figure 4 have used a fairly straightforward scaling of the visual variables using perceptual scaling recommendations where available.

From the visual impression of the maps in Figure 4 all four provide a richer representation than the standard mapping alternatives. Both attribute dimensions – the number of households and the membership value of the respective counts – are portrayed. However, the perception of each dimension and the effectiveness of each representation to convey correct evaluations and association of vagueness remains unclear until further user evaluations have been performed. Nevertheless, a few technical remarks will be presented to guide this research and future development of this technique.

### Visibility and resolution acuity

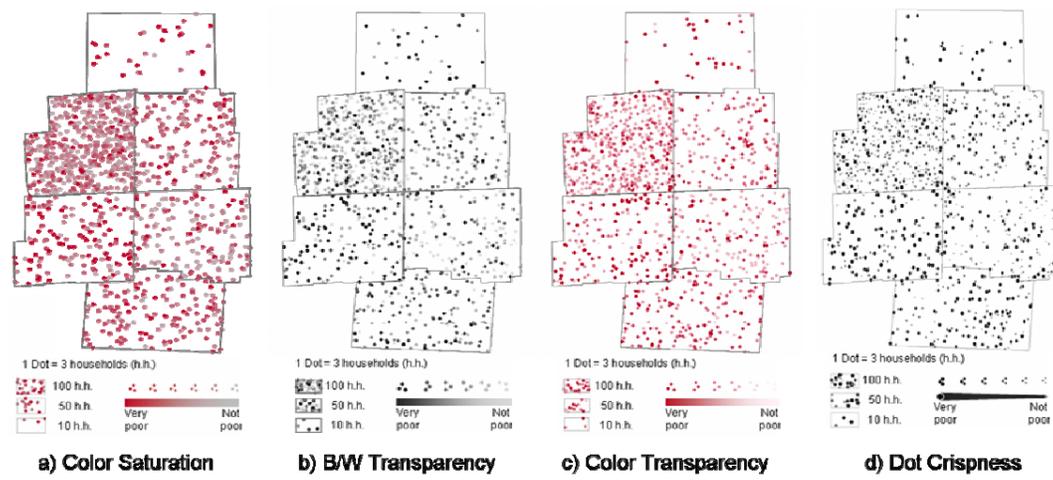


Figure 4 Maps illustrating the employment of all four visual approaches – color saturation, transparency of both color and black dots, and dot crispness – using the same hypothetical data as in Figures 1 and 3.

Acuity refers to the visibility threshold for human vision and general recommendations is that line art is not to be rendered any smaller than ~0.15mm in a regular reading situation at 40-50cm distance from a printed map (Dent 1999, p.249). The suggested dot density symbology is not only small but it also embeds crispness, color, and transparency variation, so there are potentially perceptual limits as well as screen and printer resolution issues. Dot maps commonly use dot sizes somewhat above the visual acuity, but with lower membership values, the transparency and crispness symbology can be problematic. Obviously, the symbology makes a conscious effort to approach the visibility threshold in order to achieve the effect of a symbology where objects tend to disappear when their membership values indicate that they are not part of the fuzzy set. The following remarks seek to establish an idea of some key limiting factors for the suggested symbolizations.

All approaches will be limited by printer or interface resolution. In the case of screen resolutions all color variation can be accommodated by one pixel as it can handle the mixing of colors at the sub-pixel screen element level. A square LCD pixel could then approximate a circular point down to the size of the screen resolution which today commonly go beyond 100-150 pixels/inch and thus lie near or at the recommended acuity

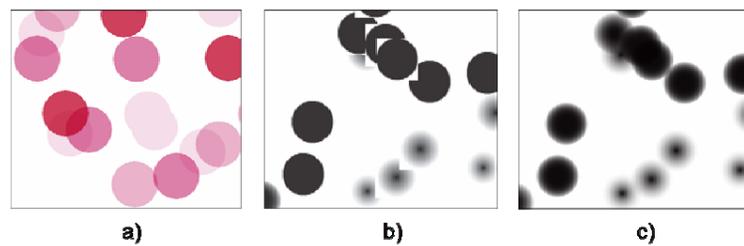


Figure 5 Close-up illustrations of issues with a) dot overlap, b) lack of support for object specific transparency, and c) ideal transparency support generated in Adobe Illustrator.

level above. For print purposes the mixing of colors necessary to produce the saturation and transparency variation will require much higher resolutions to produce solid looking colored points, but most modern printers will produce quality results for point sizes of 0.5-1.0mm. Although crispness probably is a viable approach for larger symbols, the small dot size means that whatever the lower acuity threshold is found to be, that threshold would apply to the size of the central dark spot of the rightmost symbol in Figure 3. As an effect the solid symbol to the right will have to be several times the size of the acuity threshold – at least 1.0-1.5 mm diameter – which could result in limited options to mitigate problems with overlapping dots.

#### **Dot overlaps**

General guidelines for dot value and size recommend that there is at least 2-3 dots in any mapped area unit, and that size is set so that dots in the highest density areas begin to overlap or *coalesce*. The transparency alternatives will add a complication in that overlapping points will produce a less transparent/darker tone than the dot membership values represent, see Figure 5a. Unless overlapping dot transparencies can be handled such that intersecting areas is restricted to take values within the range of overlapping object transparency values, the result can be misleading and exaggerate the degree of membership in an overcrowded area. Because of limited support for object specific transparency in the experimental implementation for this work, there is also problems with overlapping symbols that used the crispness variation. None of the symbology graphics formats in ArcMap 9.2, used in this implementation, currently support transparency (picture symbols only take BMP and EMF images) so the final result is affected by the square outlines of the picture symbology, see Figure 5b. Ideally, any overlapping symbology should transition in a seamless manner such as the other example in Figure 6c, generated in Adobe Illustrator.

#### **CONCLUSION**

Most current efforts to visualize uncertainty in geographic data have focused on error, but this is only one of at least three important types of uncertainty; error, vagueness and ambiguity. Using a synthetic data set I have illustrated the importance of vagueness in geographic data and discussed the challenges involved in adding uncertainty information to some standard types of maps. The examples demonstrated a potential to adjust the standard symbology of the dot density technique with the visual variables saturation, transparency, and crispness. My demonstration of these approaches indicated that this symbolization is capable of integrating information on quantity, pattern, and vagueness. Some design issues related to hardware resolution, symbol overlap, and object specific transparency were also highlighted and will serve as input for future studies. Work is currently underway to conduct user test of this mapping technique and other potential visual metaphors for the portrayal of vague phenomena. These studies will include the perception of symbol color and design for these and other types of maps, such as the proportional symbol map. Although similar studies exist, very little is known about e.g. how transparency and varying crispness correspond perceptually to the notion of vagueness and membership. Given the ubiquitous nature of vagueness in geographic

information this research direction should help guide future application of visualization approaches to uncertain data.

#### REFERENCES

- Anselin, L., 1999. Interactive techniques and exploratory spatial data analysis. In *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley, pp. 251-264.
- Bastin, L., Fisher, P.F. & Wood, J., 2002. Visualizing uncertainty in multi-spectral remotely sensed imagery. *Computers & Geosciences*, 28(3), 337-350.
- Burrough, P.A., MacMillan, R.A. & van Deursen, W., 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science*, 43(2), 193-210.
- Buttenfield, B. & Beard, M.K., 1994. Graphical and geographical components of data quality. In *Visualization in Geographic Information Systems*. London: John Wiley & Sons, pp. 150-157.
- Buttenfield, B.P. & Weibel, R., 1988. Visualizing the quality of cartographic data. In *Third International Geographic Information Systems Symposium (GIS/LIS 88)*, San Antonio, Texas.
- CIESIN, 2006. Where the Poor are: An Atlas of Poverty. Available at: <http://www.ciesin.columbia.edu/povmap/>.
- Coulson, M.R.C., 1990. In Praise of Dot Maps. *International Yearbook of Cartography*, 15, 51-61.
- Dahlberg, R.E., 1967. Towards the Improvement of the Dot Map. *International Yearbook of Cartography*, 7, 157-167.
- Davis, B., 2003. *Choosing a method for poverty mapping*, Rome: Food and Agriculture Organization of the United Nations. Available at: <ftp://ftp.fao.org/docrep/fao/005/y4597e/y4597e00.pdf>.
- Dent, B.D., 1999. *Cartography: thematic map design* 5th ed., William C. Brown.
- Drecki, I., 2002. Visualisation of Uncertainty in Geographic Data. In *Spatial Data Quality*. pp. 140-159.
- Fisher, P., 2000. Sorites paradox and vague geographies. *Fuzzy Sets and Systems*, 113(1), 7-18.
- Fisher, P.F., 1996. Animation of Reliability in Computer-generated Dot Maps and Elevation Models. *Cartography and Geographic Information Systems*, 23(4), 196-205.
- Fisher, P.F., 1999. Models of uncertainty in spatial data. In *Geographical Information Systems - Principles and Technical Issues*. New York: Wiley, pp. 191-205.
- Gahegan, M. et al., 2002. Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography. *Computers, Environment and Urban Systems*, 26(4), 267-292.
- Gershon, N., 1998. Visualization of an imperfect world. *Computer Graphics and Applications, IEEE*, 18(4), 43-45.

- Glasmeier, A., 2006. *An Atlas of Poverty in America: One Nation, Pulling Apart, 1960-2003*, Routledge.
- Golledge, R.G., 2002. The Nature of Geographic Knowledge. *Annals of the Association of American Geographers*, 92(1), 1-14.
- Hengl, T., 2003. Visualisation of uncertainty using the HSI colour model: computations with colours. In *Proceedings of the 7th International Conference on GeoComputation*. pp. 8-17.
- Hyman, G., Larrea, C. & Farrow, A., 2005. Methods, results and policy implications of poverty and food security mapping assessments. *Food Policy*, 30(5/6), 453-460.
- Klir, G.J. & Wierman, M.J., 1999. *Uncertainty-Based Information: Elements of Generalized Information Theory*, Springer.
- Leyk, S., Boesch, R. & Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS*, 9(3), 291-322.
- MacEachren, A.M., 1995. *How Maps Work: Representation, Visualization, and Design*, New York: Guilford Press.
- MacEachren, A.M., Brewer, C.A. & Pickle, L.W., 1998. Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, 30(9), 1547-1561.
- MacEachren, A.M. et al., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3), 139-161.
- Nelson, E.S., 2000. Designing Effective Bivariate Symbols: The Influence of Perceptual Grouping Processes. *Cartography and Geographic Information Science*, 27(4), 261-278.
- Pawling, J.W., 1973. Trend surface analysis of local relief in the southern peninsula of Michigan. *The Professional Geographer*, 25(3), 226-232.
- Plane, D.A., 1998. Fuzzy-set Migration Regions. *Geographical and Environmental Modelling*, 2(2), 141-162.
- Robinson, V.B., 2003. A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems. *Transactions in GIS*, 7(1), 3-30.
- Rogers, J.E. & Groop, R.E., 1981. Regional Portrayal With Multi-Pattern Color Dot Maps. *Cartographica*, 18(4), 51-64.
- Schultz, G.M., 1961. Using dots for traffic flow maps. *The Professional Geographer*, 13(1), 18-19.
- Ullman, E.L., 1949. Mapping the world's ocean trade: A research proposal. *The Professional Geographer*, 1(2), 19-22.
- Veregin, H., 1999. Data quality parameters. In *New Developments in Geographical Information Systems: Principles, Techniques, Management and Applications*. Wiley, pp. 177-189.
- Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8, 338-353.
- Zhang, J. & Goodchild, M.F., 2002. *Uncertainty in geographical information*, London; New York: Taylor & Francis.