International Cartographic Association
Association Cartographique Internationale

A visual approach to creating multivariate geovisualization test data

Thomas Lerch, Susanne Bleisch *

FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz/Basel, Switzerland, thomas.lerch@students.fhnw.ch, susanne.bleisch@fhnw.ch

* Corresponding author

Keywords: synthetic data generation, multivariate, spatial data, glyph visualization, usability testing

Abstract:

Testing multivariate geovisualizations, for example glyph designs, for their perceptual qualities requires suitable test data. Synthetic data can be useful for evaluating different data characteristics and their perceptibility through different designs. Real data may not contain all relevant data characteristics, for example with regard to spatial distributions or trends, which may be interesting for testing. Additionally, real data may contain a lot of noise or randomness. Even though, real data are imperative for testing design performance in realistic situations. Fuchs et al. (2017) did a systematic review on 64 glyph evaluation studies, whereof more than 60% used synthetic data. They note that for a better understanding of glyph designs more studies should evaluate glyphs with synthetic and real data. Besides visualization evaluation, a number of application areas, e.g. machine learning or software testing, generate and use synthetic data. Thus, many methods and tools for generating synthetic data exist. Typically, a mathematical function or statistical distribution is defined and random or ordered values, optionally overlaid with noise, are drawn from the defined models to build the synthetic data. However, we found it difficult to create, especially multivariate, spatial distributions of data that follow specific rules and display interactions between the data dimensions with existing tools. Thus, we designed a process that allows the intuitive 'drawing' of spatial data distributions and subsequently the derivation of multivariate data from several overlaid layers of 'drawings'.

A standard graphic software, e.g. gimp, lets us draw a data dimension in greyscale. We define that darker grey denotes larger values and lighter grey denote smaller values. The greyscale values of each pixel or of pixels at defined intervals, for example within an overlaid grid, will result in the generated data values. The different brush and gradient functions of the graphic software allow simulating different data distributions and trends (Figures 1 & 2, top). Additionally, the eraser allows clearing parts of the drawing, for example, where an imaginary wall shall limit the data distribution (Figure 1, right). A software prototype then derives data values from the greyscale images, for example for visualization of the data as glyphs (Figure 2, bottom).

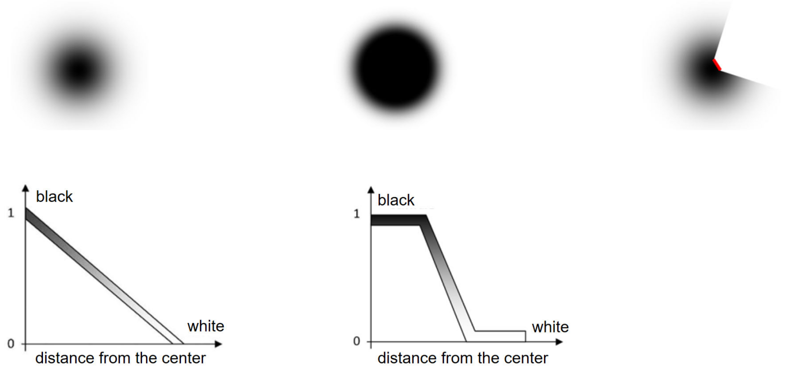

Figure 1: Data drawings based on points, left: linear decrease, middle: steep linear decrease with steps, right: linear decrease with an imaginary wall (red) limiting the distribution

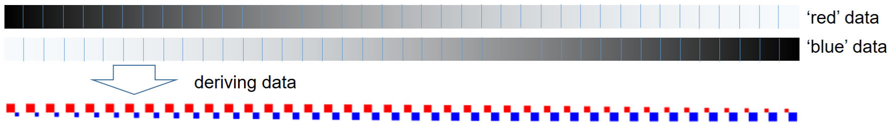

Figure 2: Two linear greyscale data drawings (top) and the, at defined intervals, derived data visualised as 2-dimensional glyphs (bottom)

Abstracts of the International Cartographic Association, 1, 2019. © Authors 2019. CC BY 4.0 License.
29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan | https://doi.org/10.5194/ica-abs-1-211-2019

Creating several data dimensions requires multiple greyscale drawings on separate layers (Figure 2, top). Our developed software prototype currently allows combining up to nine layers to derive 9-dimensional data. Masking functionality of the graphic software allows, besides erasing values and areas, limiting the data generation to certain areas, for example, along roads or for enclosed spaces. Figure 3 shows an example of a street network, duplicated four times. On each street network copy, we draw a different simulated data distribution, for example slope or accessibility, along the roads or for points on the street network using graphic software functionality. The four layers are then overlaid and, for pixels at defined intervals, the greyscale values extracted to arrive at a four-dimensional data set covering the street network. The four-dimensional data is visualised using four squares combined to glyphs (see Bleisch & Hollenstein 2017 for details on the glyph design). The data derivation prototype allows applying noise to the extracted data values in case a test setting requires it.

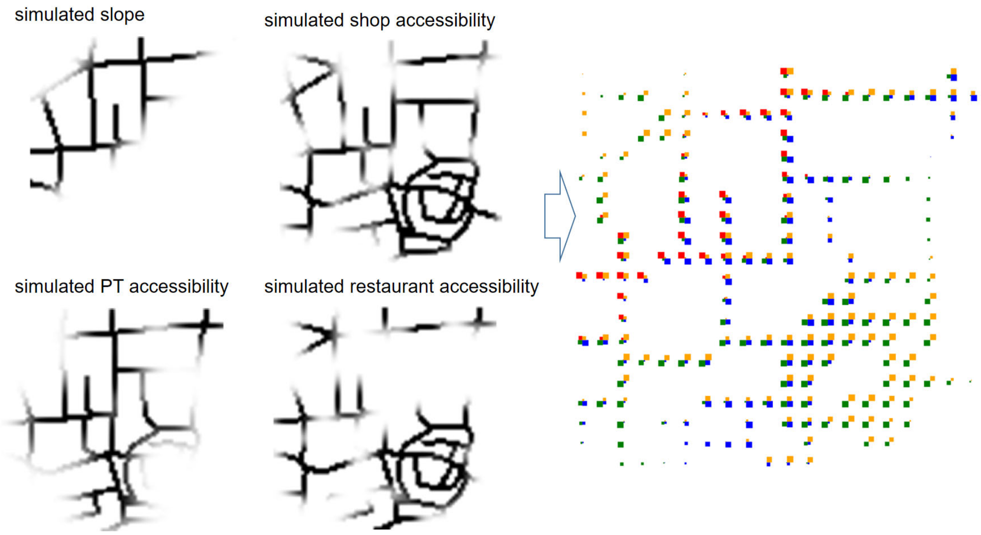
simulated slope
simulated shop accessibility
simulated PT accessibility
simulated restaurant accessibility

Figure 3: Four simulated data dimensions are drawn on the masked street network (left). The four layers are overlaid and at defined intervals the pixel values extracted to create a four-dimensional data set, visualised through glyphs (right).

The presented approach of drawing data dimensions and subsequently deriving multivariate data allows easy and intuitive synthetic data generation while still allowing the use of models through the application of, for example, linear or gaussian greyscale gradients. The biggest advantages of the process is the visual definition of the synthetic data and the simple creation of areal or masked data distributions. The visual definitions on different layers allows overlaying the different data dimensions before deriving the values and thus either creating or checking for interactions between different data dimensions. The developed software prototype then turns the layered drawings in multivariate data sets that can, for example, be used for testing of glyph-based geovisualizations.

References

Bleisch, S., & Hollenstein, D. (2017). Exploring multivariate representations of indices along linear geographic features. In *ICC 2017: Proceedings of the 2017 International Cartographic Conference*. Washington DC.

Fuchs, J., Isenberg, P., Bezerianos, A., & Keim, D. (2017). A systematic review of experimental studies on data glyphs. *IEEE Transactions on Visualization and Computer Graphics*, 23(7), 1863–1879. https://doi.org/10.1109/TVCG.2016.2549018