Proportional Symbol Mapping in R

Susumu Tanimura
Nagasaki University

Chusi Kuroiwa
University of Tokyo

Tsutomu Mizota
Nagasaki University

Abstract

Visualization of spatial data on a map aids not only in data exploration but also in communication to impart spatial conception or ideas to others. Although recent cartographic functions in R are rapidly becoming richer, proportional symbol mapping, which is one of the common mapping approaches, has not been packaged thus far. Based on the theories of proportional symbol mapping developed in cartography, the authors developed some functions for proportional symbol mapping using R, including mathematical and perceptual scaling. An example of these functions demonstrated the new expressive power and options available in R, particularly for the visualization of conceptual point data.

Keywords: visualization, cartography, thematic map, propotional symbol, R.

1. Introduction

The visualization of spatial data on a map is crucial to impart spatial conception or ideas to others or to explore spatial data. This spatial data is usually expressed on a common thematic map, such as a choropleth, proportional symbol, isarithmic, or dot map.

Each type of spatial data can be expressed using the most suitable type of thematic mapping method mentioned above for the best representation of data. Therefore it is essential to carefully select a suitable mapping method in order to draw a thematic map. However, some mapping software occasionally do not support the mapping method selected as the best choice. Sophisticated mapping software should support many types of mapping so that users can select one for the best representation of their spatial data.

Recently, cartographic packages in R including **maptools**, **mapproj**, **rgdal**, **Rmap**, and **RArcInfo** have been developed more actively and improved. This has facilitated the construction of thematic maps such as choropleth and dot maps. Despite these recent enhancements in cartographic functions in R, proportional symbol mapping, which is one of the common mapping approaches, has not been packaged thus far.
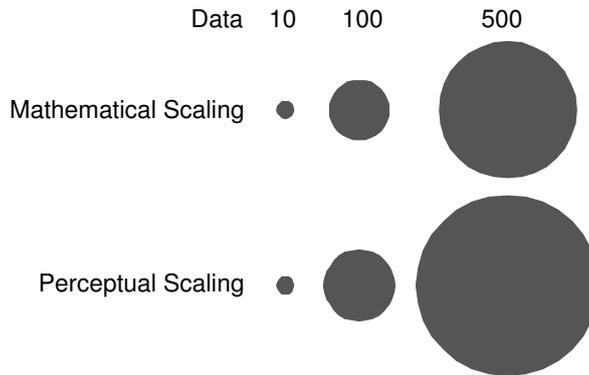
Figure 1: Contrast between mathematical and perceptual scaling

The objective of this study is to introduce proportional symbol mapping in R. We demonstrate proportional symbol mapping with new functions written in R, and discuss the feasibility and limitations of the current code.

2. Proportional symbol mapping

A proportional symbol map — one of the common thematic maps — represents spatial point data with a symbol, whose area varies in proportion to an attribute variable. The symbol used could be a circle, square, bar, sphere, cube, or a more complicated symbol such as a pictographic one.

The methodology of proportional symbol mapping has been discussed in detail by cartographers. There have been a number of related reports, and many theories have been proposed (Slocum 1999). In particular, the scaling method has been an issue. Two types of scaling techniques are widely used, namely, mathematical scaling and perceptual scaling.

In this study, we developed a function for proportional symbol mapping based on the abovementioned theories. A legend function was also developed by considering the discussion on cartography.

The type of symbol used in the function, however, was limited to a circle because the difference in the shape of the symbols is not essential for avoiding an inappropriate map expression (see discussion for more reason).

2.1. Mathematical scaling

The size of the variable in the data proportionally corresponds to the area of a point symbol. For instance, if a data value is five times another, the area of the point symbol will be five times as large. The relation is expressed as follows:

$$\frac{\pi r_i^2}{\pi r_{\max}^2} = \frac{v_i}{v_{\max}},$$

where r_i is the radius of the circle to be drawn; $r_{\max}$, radius of the largest circle on the map;

v_i, value of the variable for which the circle will be drawn; and $v_{\max}$, the maximum value of the variable.

By solving for r_i, we obtain

$$r_i \quad = \quad \left(\frac{v_i}{v_{\max}} \right)^{0.5} \times r_{\max}. \tag{1}$$

This formula was implemented in R code as the default scaling for drawing a map.

2.2. Perceptual scaling

It is well known that the perceived area of proportional symbols dose not match their mathematical area; rather, we are inclined to underestimate the area of larger symbols. As a solution to this problem, it is reasonable to modify the area of larger circles in order to match it with the perceived area.

Flanney (1971) experimentally derived a power function exponent of 0.5716 to adjust for this mismatch. For perceptual scaling, we can replace the exponent in (1) with the following approximation:

$$r_i \quad = \quad \left(\frac{v_i}{v_{\max}} \right)^{0.57} \times r_{\max}.$$

This figure is still widely cited and used for proportional symbol mapping (Slocum 1999).

Figure 1 shows an example of circles drawn with mathematical and perceptual scaling to understand the extent of the difference between them. Since perceptual scaling adjusts the area of circles in order to account for underestimation, the area of the larger circle in perceptual scaling is larger than that in mathematical scaling.

2.3. Legend design

In proportional symbol mapping, two basic legend arrangements are used: nested and linear (Figure 2). In the nested legend arrangement, a large circle includes a smaller one in sequence, while in the linear legend arrangement, the circles are aligned vertically or horizontally in the ascending or descending order. The label texts of these legends can be arranged inside or outside the circles. The advantages and disadvantages of legend arrangements have been discussed elsewhere (Slocum 1999; Slocum, McMaster, Kessler, and Howard 2005).

We included simple nested and linear arrangements in the legend function with limited flexibility. The available output of the legend function is shown in Figure 2.
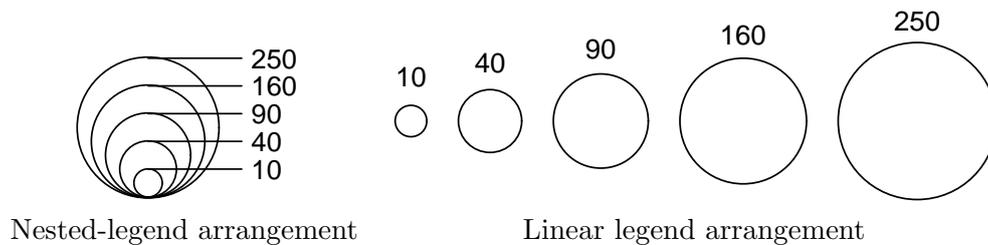

Figure 2: Nested and linear legend arrangements

3. R code

The R code of the proportional symbol mapping function (also available in the accompanying file 'ProportionalSymbolMap.R') is as follows:

```
ProportionalSymbolMap <- function(map,variable,
                                  type=c("mathematical","perceptual"),
                                  max.size=1500,
                                  symbol.fg=grey(.2), symbol.bg=grey(.5),
                                  legend.loc, legend.breaks,
                                  legend.type=c("nested","linear"),
                                  legend.cex=.6)
{
  if(missing(map)) stop("map object is missing.")
  if(!inherits(map, "Map")) stop("Map.obj must be of class Map")
  if(missing(variable)) stop("variable to be plot is missing")
  verts <- Map2points(map)[order(variable, decreasing = TRUE),]
  type <- match.arg(type)
  switch(type,
         mathematical = scale <- sqrt(variable/max(variable))*max.size,
         perceptual   = scale <- ((variable/max(variable))^0.57)*max.size)
  scale <- scale[order(variable, decreasing = TRUE)]
  symbols(verts[,1:2],circle=scale/1.1,bg=symbol.bg,fg=symbol.fg,
          inches=FALSE,add=TRUE)
  if((! missing(legend.loc)) & (! missing(legend.breaks))) {
    switch(type,
           mathematical = {
             legend.r <- sqrt(legend.breaks/max(variable))*max.size
           },
           perceptual   = {
             legend.r <- ((legend.breaks/max(variable))^0.57)*max.size
           })
    legend.type <- match.arg(legend.type)
    switch(legend.type,
           nested = {
             r <- rev(legend.r); b <- rev(legend.breaks)
             for (i in 1:length(r)) {
               symbols(legend.loc[1],legend.loc[2]+r[i],
                       circle=r[i]/1.1,inches=FALSE,add=TRUE,
                       bg=symbol.bg,fg=symbol.fg)
               lines(c(legend.loc[1],legend.loc[1]+1.2*r[1]),
                     rep(legend.loc[2]+2*r[i],2))
               text(legend.loc[1]+1.2*r[1]+par("cxy")[1]/2,
                    legend.loc[2]+2*r[i], b[i], adj=c(0,.5),
                    cex=legend.cex)
             }
           },
```

```
        linear = {
          r <- legend.r
          gap <- r[length(r)%/%2]
          s <- vector(); for (i in 1:length(r)) s[i] <- sum(r[1:i])
          x <- 2*s-r+(0:(length(r)-1))*gap + legend.loc[1]
          symbols(x, rep(legend.loc[2],length(r)), circles=r/1.1,
                  inches=FALSE, bg=symbol.bg, fg=symbol.fg, add=TRUE)
          text(x, legend.loc[2]+r+par("cxy")[2]*legend.cex, legend.breaks,
               adj=c(.5,1), cex=legend.cex)
        })
  }
}
```

The default set in this function shows mathematical scaling, a gray symbol, and the maximum radius of 1500. Prior to mapping proportional symbols, a base map is essentially required on the graphic device of R as a "new plot" termed in R. This is because a proportional symbol without the base map appears vague and may be confusing for the readers; however, in the future, the function can be expanded to be independent of the base map if the users strongly request this. The order of the plotting symbol is determined such that a circle overlays a larger circle.

When both the legend location (`legend.loc`) and the divided sizes of the legend circles (`legend.breaks`) are specified, the legend is drawn; otherwise, it is not drawn. The default legend is designed in the nested legend arrangement, while the linear legend arrangement is used when `legend.type="linear"`. Gaps between the legend circles in the linear type are fixed at the same length as the radius of the medium-sized circle in the legend.

Another problem is how to decide the breaks, i.e., whether to set equal intervals or select the most representative symbols appearing on the map. The solution to this problem is left to the user's decision rather than incorporating it as an automatic process in the function.

A sample output of `ProportionalSymbolMap()` is shown in Figure 3. From this figure, we can easily understand the distribution of child population in Nagasaki City.

4. Discussion

We developed the proportional symbol mapping function by using the R language rather than by the linkage (e.g. import and export) of external mapping software. It is important that all operations are completely controlled by R commands and limited to the R environment because such a condition yields immense benefits.

The advantage of proportional symbol mapping is that it displays attribute values for both the measured point location (true point data) and the representative point of a polygon (conceptual point data). In such case, this mapping method is superior to choropleth mapping because the latter requires the display of the classification of values, while proportional symbol mapping does not. Therefore, providing the proportional symbol mapping function is beneficial for R users because it provides an appropriate approach to visualize true or conceptual point data.

The current code supports only circles for symbolization because they have been used most
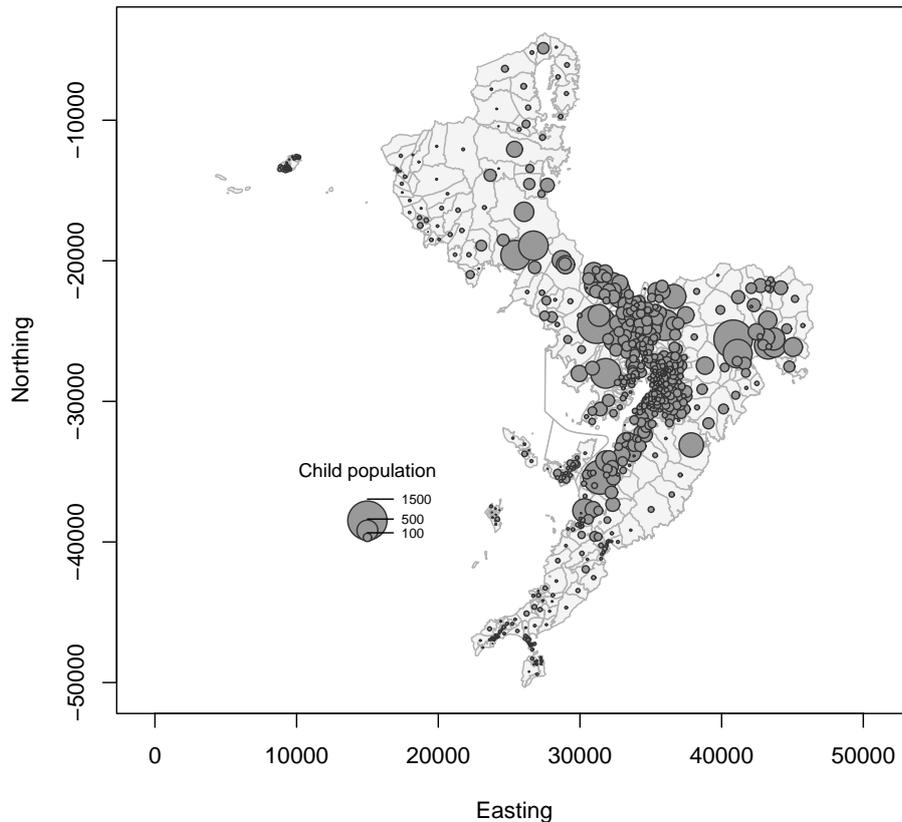
Figure 3: An example output obtained with `ProportionalSymbolMap()`. The circle denotes child population in census blocks 2000 in Nagasaki City, Japan. The symbols were scaled by perceptual scaling.

frequently, they are preferred by users, they are visually stable, and they conserve map space (Slocum 1999). However, the code can still be expanded for other geometric or pictographic symbols.

Another scaling method termed "range-graded scaling" exists in which the attribute variable to be represented is grouped into classes and symbolized with a different size of circle (Slocum 1999). In this study, we did not include this scaling method in the R code, because it is feasible only when data is grouped prior to drawing the symbols and appropriate arguments are specified. There are three issues in this method: the number of classes, method of classification, and symbol sizes, although these depend on the users' mapping policy.

The handling of symbol overlap is also a major issue in proportional symbol mapping. A small overlap with smaller symbols may result in a vague spatial pattern. On the other hand, it may be difficult to interpret individual symbols in the event of large overlap due to the stacking of larger symbols. Thus far, some solutions have been proposed; however, we have left solution to this problem to the users. With the current code, the users can decide the extent of the overlap or determine the the symbols to be used, i.e., transparent or opaque symbols.

5. Conclusion

The R function described in this paper is an initial function for proportional symbol mapping. It provides basic but sufficinet functionality for users who would like to express the attribute values of spatial point data. Consequently, R is equipped with new expressive power and more options.

Acknowledgments

References

Flanney J (1971). "The Relative Effectiveness of Some Common Graduated Point Symbols inthe Presentation of Quantitative Data." *Canadian Cartographer*, **8**(2), 96–109.

Slocum TA (1999). *Thematic Cartography and Visualization*. Prentice Hall, New Jersey.

Slocum TA, McMaster RB, Kessler FC, Howard HH (2005). *Thematic Cartography and Geographic Visualization*. Pearson Education, Inc., Upper Saddle River, NJ, 2nd edition.

Affiliation:

Susumu Tanimura
Department of Socio-environmental Medicine
Institute of Tropical Medicine
Nagasaki University
1-12-4 Sakamoto, Nagasaki, 852-8523, Japan
E-mail: stanimura-ngs@umin.ac.jp

Journal of Statistical Software
published by the American Statistical Association

http://www.jstatsoft.org/
http://www.amstat.org/

Volume 15, Issue 5
January 2006

Submitted: 2005-10-02
Accepted: 2006-01-24