

tmap: An R Package for thematic maps

by Martijn Tennekes

Abstract A thematic map is a geographical map in which statistical data are visualized. The theme refers to the statistical phenomena that is shown, such as the unemployment rate at municipal level. The best known thematic map type is the choropleth, where regions are coloured according to a statistical variable, for instance unemployment rate. Another popular thematic map type is the bubble map, in which the sizes of the bubbles are defined by a statistical variable, for instance metropolitan population. With the **tmap** package, thematic maps can be generated with great flexibility. A thematic map is created by stacking layers, for instance one for colouring municipalities, one for thick borders of higher level regions, and one for text labels. The standard work flow that is needed to create a thematic map is embedded in tmap by several convenient functions for reading, appending, and transforming spatial data.

Introduction

Visualization is key in data science. Without looking at data, it is difficult to know the data, to unveil anomalies, and, moreover, to extract valuable knowledge (Tufte, 1983). Software tools to visually explore, analyse, and present data should therefore belong to any data scientist's toolkit. The R language and its packages contain many functions to craft elegant and clarifying graphics, most notably **ggplot2** (Wickham, 2009). Although it is also possible to achieve great looking geospatial visualizations, it is often more complicated and requires the usage of several R packages together. Therefore, I introduce the **tmap** package by which thematic maps can be created in a flexible and layer-based way.

Thematic maps are geographic maps in which spatial data distributions are shown, such as population density. Although they are widely used for publication purposes, which is no surprise due to their visual appeal and recognizability, they also proved to be successful for discovery and exploration of spatial data (Friendly, 1995). One of the most illustrative examples is the dot map that physician John Snow used to locate the source of the cholera outbreak in London in 1854 (Snow, 1855). Although there are many types of thematic maps, the following five types are most commonly used:

choropleth Administrative areas, such as countries or municipalities, are filled with colors that represent a statistical variable such as population density. The usage of class intervals encourages the readability of the data values. A good starting point is to use five or seven classes which are based on quantiles (Brewer and Pickle, 2002).

proportional symbol map Data points are encoded by symbols that are scaled in proportion to a variable (Slocum et al., 2009). Bubbles are the most widely used symbols, where the corresponding maps are commonly referred to as bubble maps.

isopleth or contour map Contour lines are drawn through all points that have the same value of a quantitative variable. Commonly used in altitude and weather maps.

dot map Data points are positioned on the map by dots, like the cholera occurrences in John Snow's map. Technically, this can be seen as a specific bubble map with bubbles of constant size.

dasymetric Similar to the choropleth, except that the polygons are not predefined administrative areas, but defined by the same variable that is encoded by color. The advantage over the choropleth is that the dasymetric map portrays the underlying data distributions better (Langford and Unwin, 1994). The dasymetric map can be seen as a hybrid between the choropleth and the isopleth.

cartogram Administrative regions are distorted such that the obtained area sizes are proportional to a quantitative variable (Gastner and Newman, 2004).

The aim of the **tmap** package is to provide R users an elegant and flexible way of making thematic maps, with a minimum of code required. The syntax resembles the syntax of **ggplot2**, and follows the layered grammar of graphics (Wickham, 2010) broadly. At the time of writing, all of the described thematic map types can be created with **tmap**, except the cartogram. The implementation is not a collection of stand-alone functions, one for each thematic map type, but a coherent system in which basic layers can be stacked to create any hybrid kind of thematic map.

An illustrative example

A thematic map of the world that has been created with **tmap** is depicted in Figure 1 to illustrate the layered approach. It shows the relation between level of income per country and the distribution of

emerging metropolitan areas. While the metropolitan areas in high income countries have a annual growth rate that is less than two percent (notice that the world population increases with 2.7 percent), metropolitan areas in lower income countries grow very fast, especially in Asia but also western and middle Africa.

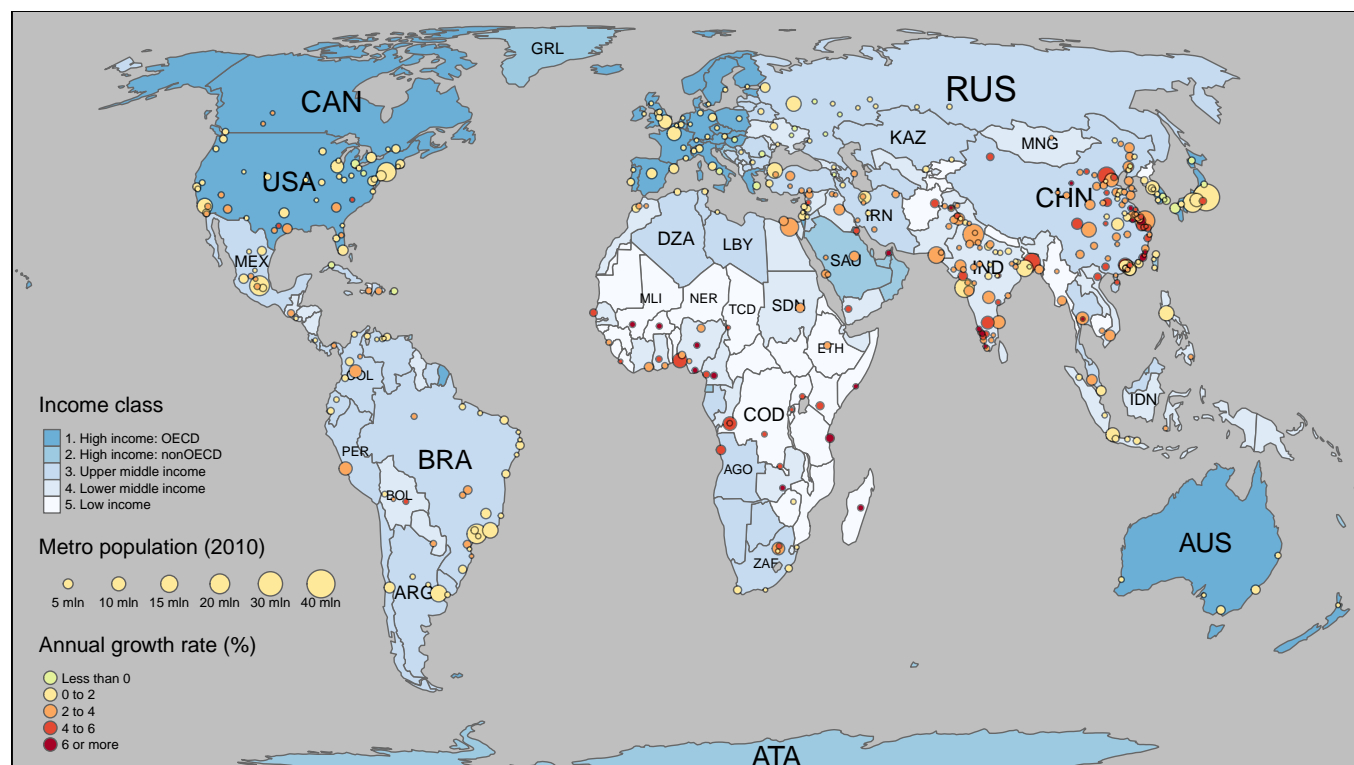


Figure 1: World map of income and urbanization.

This map is a choropleth and bubblemap in one. Although it may aesthetically look pretty and does contain a lot of information, there is one note of caution from a methodological point of view to be aware of. The contrast of yellow symbols is larger against a dark blue background than against a light blue background. In this map, the yellow colored metropolitan areas in North America attract more perceptual attention than the yellow colored metropolitan areas in Central and South America. Although this may still be a useful map for certain purposes, such as explaining the `tmap` package, it is better to have a bubble map with a uniform background to analyse the data distribution of the metropolitan areas.

The code required to create this map is the following.

```
#load spatial objects contained in tmap
data(World)
data(metro)

#derive new variable
metro$growth <- (metro$X2020 - metro$X2010) / (metro$X2010 * 10) * 100

#plot
tm_shape(World) +
  tm_fill("income_grp", palette="Blues") +
  tm_borders() +
  tm_text("iso_a3", size="AREA", scale=1.5) +
tm_shape(metro) +
  tm_bubbles("X2010", col = "growth", border.col = "black",
    border.alpha = .5, style="fixed", breaks=c(-Inf, 0, 2, 4, 6, Inf) ,
    palette="RdYlGn", contrast=1) +
tm_layout_World(title="", legend.titles=c(fill="Income class",
  bubble.size="Metro population (2010)", bubble.col="Annual growth rate (%)""))
```

The `#plot` part of the code is exemplified in detail in section 2.4. In a nutshell, two groups of layers are drawn. The first group defines a choropleth that uses the `SpatialPolygonsDataFrame` object

World. The drawing layers define the fill colors of polygons (`tm_fill`), their borders (`tm_borders`), and the supplementary text labels (`tm_text`). The second group defines a bubble map based on the `SpatialPointsDataFrame` `metro`. This group has only one layer, `tm_bubbles`, which create the bubbles.

For less complex thematic maps, there is a shortcut function called `qtm`, which stands for quick thematic map. The choropleth part of the thematic map in Figure 1 is reproduced by the following code:

```
qtm(World, fill="income_grp", text="iso_a3",
    fill.palette="-Blues",
    text.size="AREA", text.scale=1.5,
    inner.margins=c(0,-.06,.02, -.03), title="Income class")
```

It is comparable to `ggplot2`'s `qplot`; the first argument is the (spatial) data, and the next arguments are the aesthetics, in this case fill for the polygons fill color, and text, for the text labels. The remaining arguments are passed on to the underlying functions.

Related R-packages

The `tmap` package stands on the shoulders of the following six great packages.

sp (Pebesma, 2005; Roger S. Bivand, 2013) This package contains classes and methods for spatial data.

rgdal (Bivand et al., 2014) Reading and writing ESRI shape files is done by this package, as well as map projections via the PROJ.4 library (Warmerdam and Evenden, 2012).

rgeos (Bivand and Rundel, 2014) All geometric processing is carried out by this package.

classInt (Bivand, 2015) This package facilitates class intervals, which are used in choropleths of quantitative variables.

RColorBrewer (Bivand, 2015) All used color schemes have been proposed by Brewer et al. (2003) and implemented in this package.

grid This package provides all basic plotting functions.

Syntax

Another section

There will likely be several sections, perhaps including code snippets, such as:

Summary

This file is only a basic article template. For full details of *The R Journal* style and information on how to prepare your article for submission, see the [Instructions for Authors](#).

Martijn Tennekes
Statistics Netherlands
CBS-Weg 11, 6412 EX Heerlen
Netherlands
mtennekes@gmail.com

Bibliography

- R. Bivand. *classInt: Choose univariate class intervals*, 2015. URL <http://CRAN.R-project.org/package=classInt>. R package version 0.1-22. [p3]
- R. Bivand and C. Rundel. *rgeos: Interface to Geometry Engine - Open Source (GEOS)*, 2014. URL <http://CRAN.R-project.org/package=rgeos>. R package version 0.3-8. [p3]
- R. Bivand, T. Keitt, and B. Rowlingson. *rgdal: Bindings for the Geospatial Data Abstraction Library*, 2014. URL <http://CRAN.R-project.org/package=rgdal>. R package version 0.9-1. [p3]

- C. A. Brewer and L. Pickle. Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. *Annals of the Association of American Geographers*, 92(4):662–681, 2002. [p1]
- C. A. Brewer, G. W. Hatchard, and M. A. Harrower. ColorBrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30(1):5–32, 2003. [p3]
- M. Friendly. Milestones in the history of thematic cartography, statistical graphics, and data visualization. In *13TH INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS (DEXA 2002)*, AIX EN PROVENCE, pages 59–66. Press, 1995. [p1]
- M. T. Gastner and M. E. J. Newman. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20):7499–7504, 2004. [p1]
- M. Langford and D. J. Unwin. Generating and mapping population density surfaces within a geographical information system. *Cartographic Journal, The*, 31(1):21–26, June 1994. [p1]
- R. B. Pebesma, E.J. Classes and methods for spatial data in r. *R News* 5 (2), 2005. URL <http://cran.r-project.org/doc/Rnews/>. [p3]
- V. G.-R. Roger S. Bivand, Edzer Pebesma. *Applied spatial data analysis with R*. Springer, NY, second edition edition, 2013. URL <http://www.asdar-book.org/>. [p3]
- T. A. Slocum, R. B. MacMaster, F. C. Kessler, and H. H. Howard. *Thematic cartography and geovisualization, 3rd Edition*. Pearson Prentice Hall, 2009. [p1]
- J. Snow. On the mode of communication of cholera. London, 1855. [p1]
- E. R. Tufte. *The visual display of quantitative information*. Graphics Press, 1983. [p1]
- F. Warmerdam and G. Evenden. Proj.4 - cartographic projections library, version 4.8.0, 2012. URL <http://trac.osgeo.org/proj/>. [p3]
- H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL <http://had.co.nz/ggplot2/book>. [p1]
- H. Wickham. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1): 3–28, 2010. [p1]