Finnish statistics on the map in R Shiny

Authors

Pyry Kantanen, Markus Kainu, Leo Lahti

Author keywords

- Free and open-source software
- interactive visualization
- collaborative development
- R Shiny

Topics

- Historical maps and georeferencing
- Labs notebooks
- workflows and infrastructure

Abstract

Visualization of statistical data has long been recognized as an important part of understanding data, interpreting results, and communicating findings and results of scientific research. This is more true than ever with advances in technology related to authoring, publishing, and consumption of scientific papers, as well as societies in general becoming increasingly reliant on visual communication channels. (Midway, 2020.) Visualising statistical data in the form of choropleth maps – or thematic maps – represents an important subset of data visualization. Statistical data projected on different administrative and geographic boundaries, in other words, geospatial data visualisation, communicates geographic facts and provides a starting point for spatial reasoning (Barrozo, et al., 2016).

Choropleth mapping features are accessible in various software solutions, for example Microsoft Power BI, open-source QGIS and open-source Oskari web mapping framework. In this paper, we argue that the R ecosystem, which is traditionally perceived as an open-source scripting language for statistics and data science, offers a balanced and relatively easy-to-use solution to most common geospatial data visualization needs, both as high-quality renders for scientific papers and other traditional publications as well as interactive web apps with R Shiny. In addition, the R ecosystem provides the flexibility of the open-source model of development, enabling more experienced users to implement new features rapidly.

The geofi R package, developed as part of the rOpenGov community project for open government data R packages, provides tools to facilitate access to Finnish geospatial datasets in a fast and user-friendly manner. The package, alongside other rOpenGov packages such as pxweb (Magnusson, et al., 2021), sotkanet (Lahti, et al., 2022) and eurostat (Lahti, et al., 2017) developed with the Finnish and European open data infrastructure in mind, make it possible to create scripted, automated solutions for downloading and visualising statistics over different administrative divisions of Finland from the national level to several different flavours of subnational divisions, from regions and major regions to wellbeing service counties and working areas.

The ease of linking geospatial information from the geofi package to statistical data downloaded with pxweb and sotkaweb packages is demonstrated by the R Shiny app "Tilastot kartalle" (Engl. "Statistics on the map"). The app, in its current form only in English, is aimed to make creating and saving simple choropleth maps and bar plots as convenient as possible for end-users of all skill levels. For users who are already acquainted with using R or users who are willing to learn the syntax and do their own modifications, the app prints out code for drawing the visualizations on their local R installation. This code transparency enables users to break free of the pre-made choices and limitations of the Shiny user interface, experiment with the code and possibly even improve it by implementing new data sources and new maps. The Shiny application, like any other Shiny application, can easily be run in any graphical desktop environment that can run a recent version of R.

For future development, there we recognize two main challenges. First, the boundaries of existing administrative areas are not drawn with explorative data analysis in mind and therefore using them as a starting point for thematic mapping can lead to misleading results. For example, administrative areas can vary greatly in population density and size, making large territories with sparse populations stand out in contrast to smaller, more densely populated areas (Barrozo, et al., 2016). The geofi package could include options to augment the maps by adding 3D extrusions or deformed cartograms to illustrate the relative weight of different areas (see Besançon, et al., 2020).

Second, public data sources currently used offer mainly statistical datasets and geospatial data from the last 15 years, making historical choropleth maps impossible to make. As a collaborative software development project, we have encouraged our users to give feedback and participate in the development of the software. In the future, these collaborative practices should also include hackathons and mapathons for georeferencing historic maps and creating new packages for accessing historical datasets that are not part of official statistics, for example Finnish church records for historical population data.

References

Barrozo, L. V., Pérez-Machado, R. R. & Small, C. C.-M. W., 2016. Changing spatial perception: dasymetric mapping to improve analysis of health outcomes in a megacity. Journal of Maps, 12(5), p. 1242–1247.

Besançon, L., Cooper, M., Ynnerman, A. & Vernier, F., 2020. An Evaluation of Visualization Methods for Population Statistics Based on Choropleth Maps. arXiv preprint.

Lahti, L. et al., 2022. sotkanet: Sotkanet Open Data Access and Analysis. R package version 0.9.79. Available at: https://github.com/rOpenGov/sotkanet

Lahti, L., Huovari, J., Kainu, M. & Biecek, P., 2017. Retrieval and Analysis of Eurostat Open Data with the eurostat Package. The R Journal, 9(1), p. 385–392.

Magnusson, M., Kainu, M., Huovari, J. & Lahti, L., 2021. pxweb: R tools for PX-WEB API. R package version 0.12.0. Available at: https://github.com/rOpenGov/pxweb

Midway, S. R., 2020. Principles of Effective Data Visualization. Patterns, 1(9).