rmap: An R package to plot and compare tabular data on customizable maps across scenarios and time

Zarrar Khan[1], **Mengqi Zhao**[2], **Chris R. Vernon**[1], **and Thomas Wild**[1, 2]

1 Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA **2** Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA

DOI: 10.21105/joss.03972

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Submitted: 30 November 2021
Published: 01 December 2021

Summary

`rmap` is an R package that allows users to easily plot tabular data (CSV or R data frames) on maps without any Geographic Information Systems (GIS) knowledge. Maps produced by `rmap` are `ggplot` objects and thus capitalize on the flexibility and advancements of the `ggplot2` package (Wickham, 2011) and all elements of each map are thus fully customizable. Additionally `rmap` automatically detects and produces comparison maps if the data has multiple scenarios or time periods as well as animations for time series data. Advanced users can load their own shapefiles if desired. `rmap` comes with a range of pre-built color palettes but users can also provide any `R` color palette or create their own as needed. Four different legend types are available to highlight different kinds of data distributions. The input spatial data can be both gridded or polygon data. `rmap` is desgined in particular for comparing spatial data across scenarios and time periods and comes preloaded with standard country, state, and basin maps as well as custom maps compatible with the Global Change Analysis Model (GCAM) spatial boundaries (Calvin et al., 2019). `rmap` has a growing number of users and its products have been used in multiple multisector dynamics publications (Khan et al., 2021; Wild, Khan, Clarke, et al., 2021; Wild, Khan, Zhao, et al., 2021) as well as a required dependency in other R packages such as `rfasst` (Sampedro, 2021) and `metis` (Khan et al., 2020). `rmap's` automatic processing of tabular data using pre-built map selection, difference map calculations, faceting, and animations offers unique functionality which makes it a powerful and yet simple tool for users looking to explore multi-sector, multi-scenario data across space and time.

Statement of need

`rmap` is meant to help users having limited to no GIS knowledge use R for spatial visualization of tabular spatial data. `rmap` is not meant to be a replacement for spatial manipulation or cartographic software but focuses on the simple plotting of polygon and gridded data for spatio-temporal visualization of tabular data with a focus on comparing across scenarios and time periods. Several existing R packages such as tmap (Tennekes, 2018), cartography (Giraud & Lambert, 2016), rworldmap (South, 2011), GISTools (Brunsdon et al., 2015), choroplethr (Lamstein & Johnson, 2020), sp (E. Pebesma & Bivand, 2005) and sf (E. J. Pebesma, 2018), have been developed to conduct spatial visualization and analytics in R without depending on external software such as ArcGIS (ESRI, 2020), GRASS (GRASS Development Team, 2020) or QGIS (QGIS Development Team, 2021). `rmap` enhances the following key capabilities which are limited in these existing packages:

1. **Pre-built maps**: Existing packages come with only a few examples of built-in maps as package data. `rmap` comes with a growing collection of country, state, river basin, as well as other customized maps that are added into the package data based on user needs and requests. While built-in maps increase the size of the package, having direct access to these allows for automated searching and quick deployment of relevant shapefiles without the need to download any data. A list of pre-built maps in `rmap` can be found in the Built-in Maps section of the user guide.
2. **Direct data table to map**: Existing packages are not able to plot a map directly given only a simple data frame or a CSV file as an input. `rmap` has an automatic `map_find` function that searches for the appropriate built-in map based on the regions provided in a `subRegion` column and values in a `value` column. The sub-regions in the `subRegion` column must be one of the sub-regions in the existing set of `rmap` built-in maps. This truly frees users from the need for any other spatial data needs and they can simply `map()` their own data tables directly.
3. **Difference maps**: Existing packages do not produce difference maps to compare across scenarios or time periods. `rmap` provides this functionality by automatically recognizing multiple scenarios and time periods to produce difference maps across these dimensions. An important aspect of spatial data is exploring the difference between two scenarios or time periods and `rmap` makes this a seamless process.
4. **Post-process customization**: Existing packages do not produce output objects that can be saved and then customized. Customization of the maps in existing packages is limited to package specific functionality and arguments. `rmap` produces `ggplot` objects in which every element (axis, grids, titles, colors, line widths, facets) can all be customized after the map has been produced. This allows users to capitalize on existing knowledge of the widely used `ggplot2` package and its arguments.

Functionality

A detailed User Guide walks users step-by-step through all the available functionality of `rmap`. A simpler Cheatsheet is also provided to help users remember some of the key functionality in a single sheet. The following few simple examples demonstrate the simplicity of using `rmap`.

Basic map

```
library(rmap)

data = data.frame(subRegion = c("China","India","Pakistan","Iran","Afghanistan"),
                  value = c(5,10,15,34,2))
map(data)
```

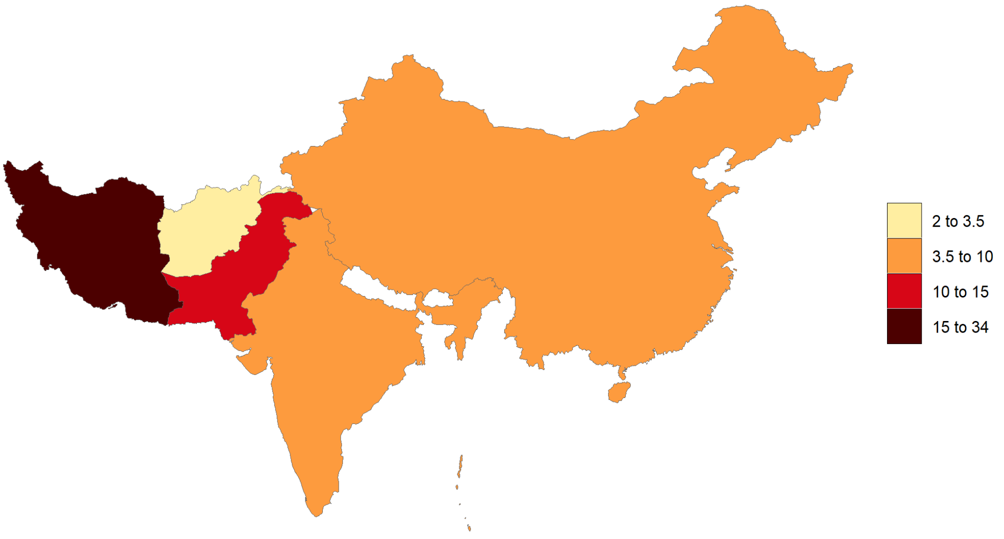

Figure 1: Basic map plotted on data provided. `rmap` automatically chooses a country map based on the names in the 'subRegion' column in the data table.

Map with labels and underlayer

```
library(rmap)

data = data.frame(subRegion = c("CA","FL","ID","MO","TX","WY"),
                  value = c(5,10,15,34,2,7))
map(data,
    underLayer = mapUS52Compact,
    crop_to_underLayer = T,
    labels = T)
```

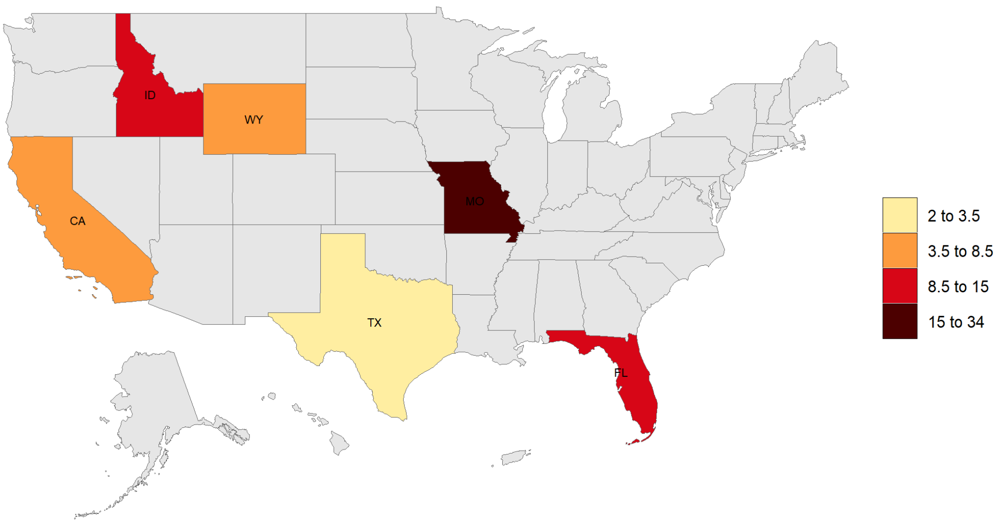

Figure 2: Map with chosen underLayer and labels for data provided. `rmap` automatically chooses the U.S. state map based on the names in the 'subRegion' column in the data table.

Compare Scenarios

```
library(rmap)

data = data.frame(subRegion = c("Spain","Germany","Austria","Greece","Italy",
                                "Spain","Germany","Austria","Greece","Italy",
                                "Spain","Germany","Austria","Greece","Italy"),
                  value = c(5,10,15,34,2,
                            15,50,34,50,20,
                            1,2,7,13,5),
                  scenario = c("scen1","scen1","scen1","scen1","scen1",
                               "scen2","scen2","scen2","scen2","scen2",
                               "scen3","scen3","scen3","scen3","scen3"))
map(data,
    underLayer = "mapCountries",
    scenRef = "scen1",
    background = T)
```

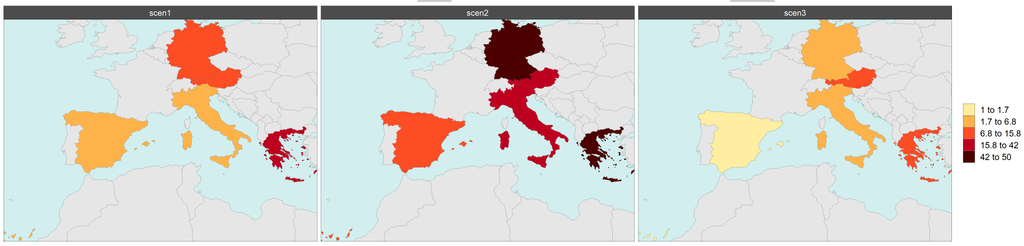

Figure 3: Maps showing absolute values for all three scenarios in data table.

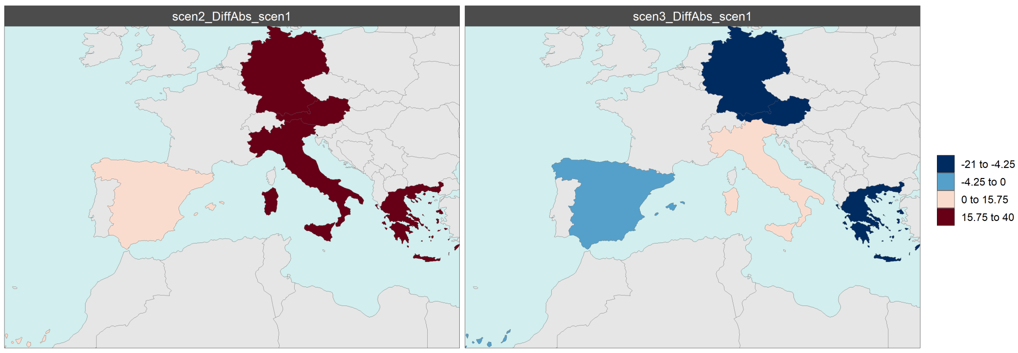

Figure 4: Maps showing absolute difference between chosen reference scenario ("scen1") and remaining scenarios ("scen2" and "scen3") in data table.

Acknowledgements

This research was supported by the US Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. The views and opinions expressed in this paper are those of the authors alone.

References

Brunsdon, C., Chen, H., & Brunsdon, M. C. (2015). Package 'GISTools.' *CRAN Repository*.

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Vittorio, A. D., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., … Wise, M. (2019). GCAM v5.1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, *12*(2), 677–698. https://doi.org/10.5194/gmd-12-677-2019

ESRI. (2020). *ArcGIS*. http://www.esri.com/software/arcgis

Giraud, T., & Lambert, N. (2016). Cartography: Create and Integrate Maps in your R Workflow. *Journal of Open Source Software*, *1*(4), 54. https://doi.org/10.21105/joss.00054

GRASS Development Team. (2020). *GRASS GIS graphical user interface*. grass.osgeo.org

Khan, Z., Wild, T. B., Iyer, G., Hejazi, M., & Vernon, C. R. (2021). The future evolution of energy-water-agriculture interconnectivity across the US. *Environmental Research Letters*, *16*(6), 065010. https://doi.org/10.1088/1748-9326/ac046c

Khan, Z., Wild, T. B., Vernon, C. R., Miller, A., Hejazi, M. I., Clarke, L. E., Miralles-Wilhelm, F. R., Muñoz Castillo, R., Moreda, F., Lacal Bereslawski, J. C., & others. (2020). Metis–a tool to harmonize and analyze multi-sectoral data and linkages at variable spatial scales. *Journal of Open Research Software*, *8*(PNNL-SA-146642). https://doi.org/10.5334/jors.292

Lamstein, A., & Johnson, B. (2020). *Choroplethr: Simplify the Creation of Choropleth Maps in R*. https://cran.r-project.org/web/packages/choroplethr/index.html

Pebesma, E., & Bivand, R. S. (2005). S classes and methods for spatial data: The sp package. *R News*, *5*(2), 9–13. https://github.com/edzer/sp

Pebesma, E. J. (2018). Simple features for R: Standardized support for spatial vector data. *R J.*, *10*(1), 439. https://github.com/r-spatial/sf

QGIS Development Team. (2021). *QGIS*. qgis.org/en/site/

Sampedro, J. (2021). *Rfasst: An R tool to estimate air pollution impacts on health and agriculture. Version 1.1*. https://doi.org/10.5281/zenodo.4960636

South, A. (2011). Rworldmap: A new R package for mapping global data. *R Journal*, *3*(1). https://doi.org/10.32614/RJ-2011-006

Tennekes, M. (2018). Tmap: Thematic Maps in R. *Journal of Statistical Software*, *84*(1), 1–39. https://github.com/r-tmap/tmap

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

Wild, T. B., Khan, Z., Clarke, L., Hejazi, M., Bereslawski, J. L., Suriano, M., Roberts, P., Casado, J., Miralles-Wilhelm, F., Gavino-Novillo, M., Muñoz-Castillo, R., Moreda, F., Zhao, M., Yarlagadda, B., Lamontagne, J., & Birnbaum, A. (2021). Integrated energy-water-land nexus planning in the Colorado River Basin (Argentina). *Regional Environmental Change*, *21*(3), 62. https://doi.org/10.1007/s10113-021-01775-1

Wild, T. B., Khan, Z., Zhao, M., Suriano, M., Bereslawski, J. L., Roberts, P., Casado, J., Gaviño-Novillo, M., Clarke, L., & Hejazi, M. (2021). The Implications of Global Change for the Co-Evolution of Argentina's Integrated Energy-Water-Land Systems. *Earth's Future*, *9*(8), e2020EF001970. https://doi.org/0.1029/2020EF001970