The Journal of Open Source Software

Using the tidyverse with terra objects: the tidyterra package

Diego Hernangómez [1]

1 Independent Researcher

DOI: 10.21105/joss.05751

Software

- Review
- Repository
- Archive

Editor: Jayaram Hariharan

Reviewers:

- @mikejohnson51
- @Yingjie4Science

Submitted: 19 July 2023
Published: 10 November 2023

Summary

`tidyterra` is an R (R Core Team, 2023) package that allows manipulation of spatial data objects as provided by the `terra` package (Hijmans, 2023), using the verbs of the packages included in the tidyverse (Wickham et al., 2019), such as `dplyr` (Wickham, François, et al., 2023), `tidyr` (Wickham, Vaughan, et al., 2023), or `tibble` (Müller & Wickham, 2023). This addition enables users that are already familiar with the tidyverse to approach spatial data manipulation and analysis more easily and much faster.

Furthermore, `tidyterra` extends the functionality of the `ggplot2` package (Wickham, 2016) by providing additional geoms and stats [1] like `geom_spatraster()` and `geom_spatvector()`, as well as carefully chosen scales and color palettes specifically designed for map production.

`tidyterra` can manipulate the following classes of `terra` objects:

1. `SpatVector` objects, which represent vector data such as points, lines, or polygon geometries.
2. `SpatRaster` objects, which represent raster data in the form of a grid consisting of equally sized rectangles. Each rectangle can contain one or more values.

The first stable version of `tidyterra` was included on CRAN on April 24, 2022, and has been actively used by other packages (such as `ebvcuve` (Quoss et al., 2021), `biomod2` (Thuiller et al., 2023), `inlabru` (Bachl et al., 2019), `RCzechia` (Lacko, 2023), and `sparrpowR` (Buller et al., 2021)) and cited in academic research and publications (Bahlburg et al. (2023), Moraga (2023), Leonardi et al. (2023), Meister et al. (2023)) ever since.

Statement of need

The tidyverse is a compilation of R packages that share an underlying design philosophy, grammar, and data structures. The packages within the tidyverse are widely used by R users for tidying, transforming, and visualizing data.

The tidyverse is designed to work with tidy data (*"every column is a variable, every row is an observation, every cell is a single value"*), represented in the form of data frames or tibbles. However, it is possible to extend the functionality of tidyverse packages to work with new R object classes by registering the corresponding S3 methods (Wickham, 2019). This means that `dplyr::mutate()` can be adapted to work with any object of class `foo` by creating the corresponding S3 method `mutate.foo()`.

While other popular packages designed for spatial data handling, such as `sf` (Pebesma, 2018) or `stars` (Pebesma & Bivand, 2023), already provide integration with the tidyverse as part of

[1] The term "geoms" refers to geometric objects, and "stats" refers to statistical transformations, following the naming conventions of `ggplot2`

their infrastructure, `terra` objects lack this integration natively. Although `terra` offers a wide set of functions for transforming and visualizing `SpatRaster` and `SpatVector` objects, some users who are not familiar with this package would need to make an additional effort to learn that syntax. This may imply an additional challenge during their initial steps in the field of spatial analysis.

The `tidyterra` package was developed to address this integration gap. By providing the corresponding S3 methods, data analysts can apply the same syntax and functions they are already familiar with for rectangular data to the objects provided by `terra`. This enables users who are not familiar with spatial data analysis to approach this area more easily.

In addition, `tidyterra` also offers functions for plotting `terra` objects using the `ggplot2` syntax. Although packages like `rasterVis` (Perpiñán & Hijmans, 2023) and `ggspatial` (Dunnington, 2023) already allow the representation of `SpatRaster` objects via `ggplot2`, `tidyterra` functions provide additional support for advanced mapping. This support includes the integration of faceted maps, contours, and the automatic conversion of spatial layers to the same CRS[2] via `ggplot2::coord_sf()`. Furthermore, `tidyterra` also provides support for `SpatVector` objects, similar to the native support of `sf` objects in the `ggplot2` package.

Lastly, `tidyterra` provides a collection of color palettes specifically designed for representing spatial phenomena (Lindsay, 2018). Additionally, it implements the cross-blended hypsometric tints described by Patterson & Jenny (2011).

A note on performance

The development philosophy of `tidyterra` consists on adapting `terra` objects to data frame-like structures by performing different data transformations, that ultimately may impact in the performance of the package.

When manipulating large raster files (i.e. more than 10.000.000 cells), it is recommended to use the native `terra` syntax, that is specifically designed for handling this type of files. In the case of plotting, the default behavior of the geoms provided is to resample `SpatRaster` that presents more than 500.000 cells to speed up the process (as the `terra::plot()` does), however this upper limit can be modified using the `maxcell` parameter of the geom function.

Note also that when possible, the help page of each function of `tidyterra` references its equivalent in `terra`.

Example of use

`tidyterra` is available on CRAN, so it can be easily installed using the following commands in R:

```
install.packages("tidyterra")
```

The latest developing version is hosted in GitHub and can be installed using the following command in R:

```
remotes::install_github("dieghernan/tidyterra")
```

The following example demonstrates how to manipulate a `SpatRaster` object using the `dplyr` syntax. Additionally, it illustrates how to seamlessly plot a `SpatRaster` object with `ggplot2` using the `geom_spatraster()` function:

```
library(tidyterra)
library(tidyverse) # Load all the packages of tidyverse at once
library(scales) # Additional library for labels
```

[2] CRS; Coordinate reference system

```r
# Temperatures in Castille and Leon (selected months)
rastertemp <- terra::rast(system.file("extdata/cyl_temp.tif",
  package = "tidyterra"
))

# Rename with the tidyverse
rastertemp <- rastertemp %>%
  rename(April = tavg_04, May = tavg_05, June = tavg_06)


# Plot with facets
ggplot() +
  geom_spatraster(data = rastertemp) +
  facet_wrap(~lyr, ncol = 2) +
  scale_fill_whitebox_c(
    palette = "muted",
    labels = label_number(suffix = "º"),
    n.breaks = 12,
    guide = guide_legend(reverse = TRUE)
  ) +
  labs(
    fill = "",
    title = "Average temperature in Castille and Leon (Spain)",
    subtitle = "Months of April, May and June"
  )
```

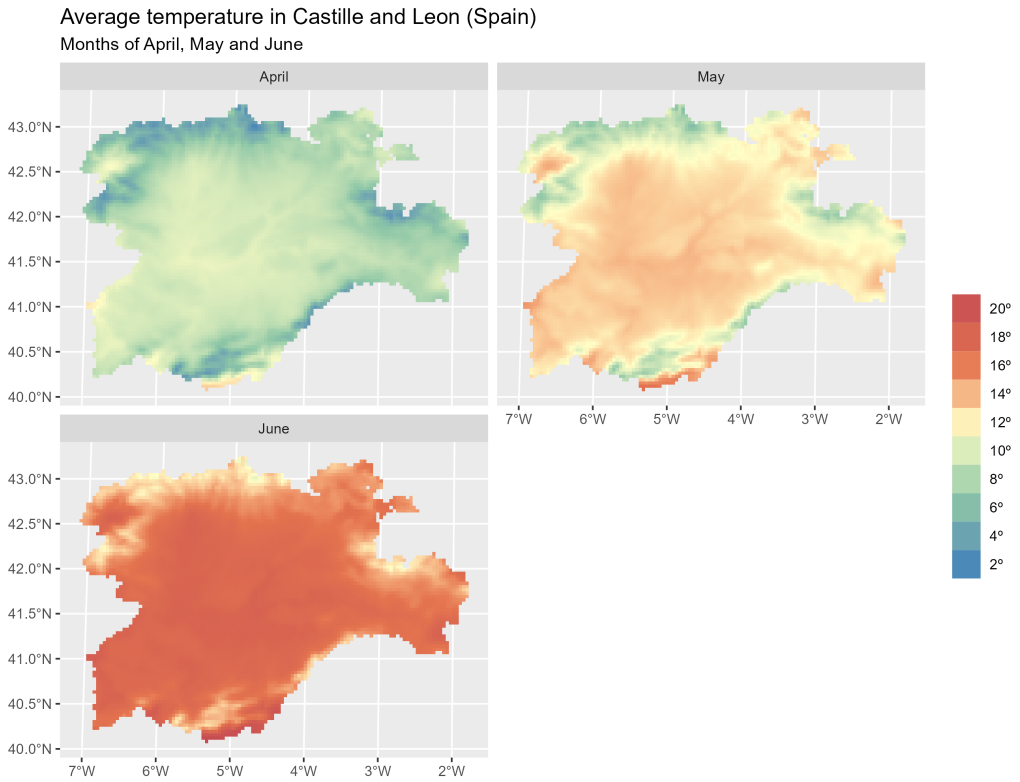

Figure 1: Faceted map with multi-layer raster file.

In the following example we combine a common `dplyr` workflow (`mutate` + `select`) and we plot the result. In this case the plot is a contour plot of the original `SpatRaster` using `geom_spatraster_contour_filled()`and it also includes an overlay of a `SpatVector` for reference:

```
# Compute the variation between April and June and apply a different palette
incr_temp <- rastertemp %>%
  mutate(var = June - April) %>%
  select(Variation = var)

# Overlay an SpatVector
cyl_vect <- terra::vect(system.file("extdata/cyl.gpkg",
  package = "tidyterra"
))

# Contour map with overlay
ggplot() +
  geom_spatraster_contour_filled(data = incr_temp) +
  geom_spatvector(data = cyl_vect, fill = NA) +
  scale_fill_whitebox_d(palette = "bl_yl_rd") +
  theme_grey() +
  labs(
    fill = "º Celsius",
    title = "Variation of temperature in Castille and Leon (Spain)",
    subtitle = "Difference between April and June"
  )
```

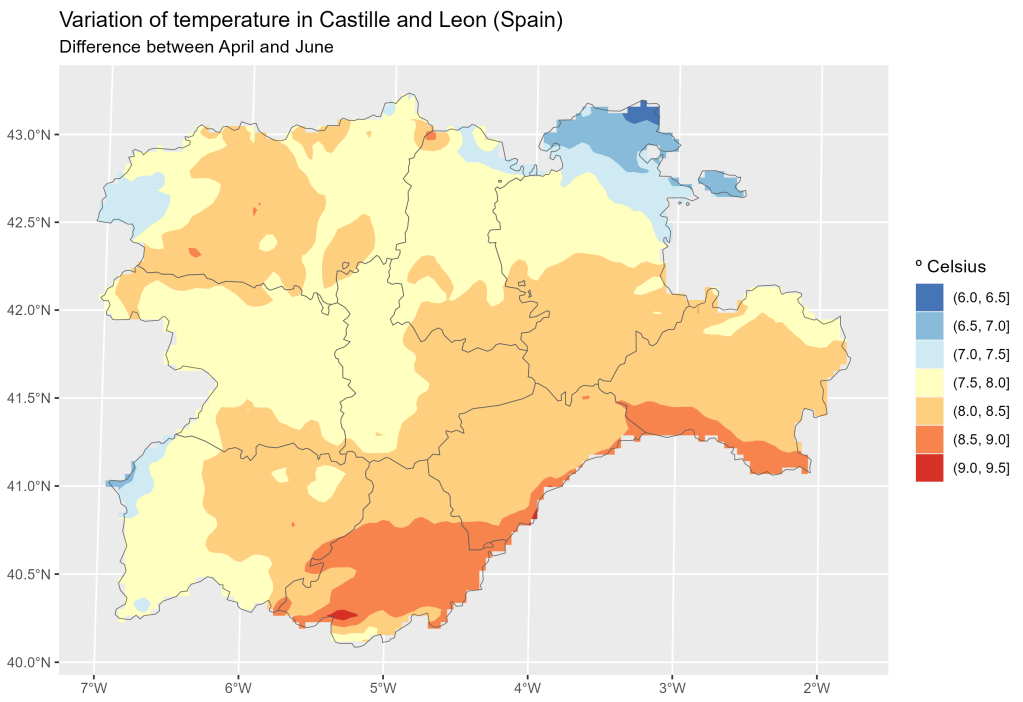

Figure 2: Contour map of temperature variation with a `SpatVector` overlay.

Additional materials

The package includes extensive documentation available online at https://dieghernan.github.io/tidyterra/ including:

- Details on each function, including (if possible) the equivalent `terra` function, in case users prefer to include those on their workflows.
- Working examples on the use of the functions and creation of plots.
- Additional articles and vignettes, as well as a complete demo of the different color palettes included on the package (see Palettes).

Acknowledgements

I would like to thank Robert J. Hijmans for his advice and support in adapting some of the methods, as well as for the suggestions that helped us improve the functionalities of the package. I am also thankful to Dewey Dunnington, Brent Thorne, and the rest of contributors of the `ggspatial` package, which served as a key reference during the initial stages of the development of `tidyterra`.

`tidyterra` also incorporates some pieces of code adapted from `ggplot2` for computing contours, which relies on the package `isoband` (Wickham et al., 2022) developed by Claus O. Wilke.

References

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*, 760–766. https://doi.org/10.1111/2041-210X.13168

Bahlburg, D., Thorpe, S. E., Meyer, B., Berger, U., & Murphy, E. J. (2023). An intercomparison of models predicting growth of Antarctic krill (Euphausia superba): The importance of recognizing model specificity. *PLOS ONE*, *18*(7), 1–29. https://doi.org/10.1371/journal.pone.0286036

Buller, I. D., Brown, D. W., Myers, T. A., Jones, R. R., & Machiela, M. J. (2021). sparrpowR: A flexible R package to estimate statistical power to identify spatial clustering of two groups and its application. *International Journal of Health Geographics*, *20*(1), 1–7. https://doi.org/10.1186/s12942-021-00267-z

Dunnington, D. (2023). *ggspatial: Spatial data framework for ggplot2*. https://CRAN.R-project.org/package=ggspatial

Hijmans, R. J. (2023). *terra: Spatial data analysis*. https://rspatial.org/

Lacko, J. (2023). RCzechia: Spatial objects of the Czech Republic. *Journal of Open Source Software*, *8*(83). https://doi.org/10.21105/joss.05082

Leonardi, M., Colucci, M., & Manica, A. (2023). tidysdm: Leveraging the flexibility of tidymodels for species distribution modelling in R. *bioRxiv*. https://doi.org/10.1101/2023.07.24.550358

Lindsay, J. (2018). *Whitebox-tools*. GitHub repository. https://github.com/jblindsay/whitebox-tools

Meister, N., Langbehn, T. J., Varpe, Ø., & Jørgensen, C. (2023). Blue mussels in western Norway have vanished where in reach of crawling predators. *Marine Ecology Progress Series*, *721*, 85–101. https://doi.org/10.3354/meps14416

Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. Chapman & Hall/CRC. https://www.paulamoraga.com/book-spatial/

Müller, K., & Wickham, H. (2023). *tibble: Simple data frames* (Version 3.2.1). https://tibble.tidyverse.org/

Patterson, T., & Jenny, B. (2011). The development and rationale of cross-blended hypsometric tints. *Cartographic Perspectives*, *69*, 31–46. https://doi.org/10.14714/CP69.20

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With applications in R* (p. 352). Chapman; Hall/CRC. https://doi.org/10.1201/9780429459016

Perpiñán, O., & Hijmans, R. (2023). *rasterVis*. https://oscarperpinan.github.io/rastervis/

Quoss, L., Fernandez, N., Langer, C., Valdez, J., & Pereira, H. M. (2021). *ebvcube: Working with netCDF for essential biodiversity variables*. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. https://github.com/LuiseQuoss/ebvcube

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., & Patin, R. (2023). *biomod2: Ensemble platform for species distribution modeling*.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4

Wickham, H. (2019). S3. In *Advanced R* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781351201315

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.2). https://dplyr.tidyverse.org

Wickham, H., Vaughan, D., & Girlich, M. (2023). *tidyr: Tidy messy data* (Version 1.3.0). https://tidyr.tidyverse.org

Wickham, H., Wilke, C. O., & Pedersen, T. L. (2022). *isoband: Generate isolines and isobands from regularly spaced elevation grids*. https://CRAN.R-project.org/package=isoband