

ceylon: An R package for plotting the maps of Sri Lanka

26 December 2023

Summary

The rapid evolution in the fields of computer science, data science, and artificial intelligence has significantly transformed the utilisation of data for decision-making. Data visualisation plays a critical role in any work that involves data. Visualising data on maps is frequently encountered in many fields. Visualising data on maps not only transforms raw data into visually comprehensible representations but also converts complex spatial information into simple, understandable form. Locating the data files necessary for map creation can be a challenging task. Establishing a centralised repository can alleviate the challenging task of finding shape files, allowing users to efficiently discover geographic data. The `ceylon` (Talagala 2023) R package is designed to make simple feature data related to Sri Lanka’s administrative boundaries and rivers and streams accessible for a diverse range of R (R Core Team 2023) users. With straightforward functionalities, this package allows users to quickly plot and explore administrative boundaries and rivers and streams in Sri Lanka.

Statement of Need

The `ceylon` R package conveniently packages shape files corresponding to the geographic features of Sri Lanka, enhancing user friendliness for seamless integration and analysis. This allows for minimising the time spent on data searching and cleaning efforts. Hence, the package `ceylon` stands out as a catalyst for research efficiency. Furthermore, the package supports research reproducibility, allowing others to independently verify and build upon the work that utilised the data in this package. The data format easily integrates with tidyverse packages (Wickham et al. 2019), fostering a smooth workflow.

Datasets available in the package

The data were retrieved from the “Humanitarian Data Exchange (HDX)” (n.d.). The Humanitarian Data Exchange is a platform that facilitates the sharing and collaboration of humanitarian data. The coordinate reference system (CRS) for the data is “Projected CRS: SLD99 / Sri Lanka Grid 1999 (CRS code 5234).” Table 1 provides a description of datasets.

dataset	description	data source
<code>sf_sl_0</code>	country boundary	https://data.humdata.org/
<code>province</code>	province boundaries	https://data.humdata.org/
<code>district</code>	district boundaries	https://data.humdata.org/
<code>sf_sl_3</code>	divisional secretariat boundaries	https://data.humdata.org/
<code>rivers</code>	Sri Lanka rivers and streams shapefiles	https://data.humdata.org/

Table 1: Description of data in the package

Usage

ceylon is available on GitHub, and can be installed and loaded into the R session using:

```
install.packages("devtools")
devtools::install_github("thiyanagt/ceylon")
library(ceylon)
```

The additional packages required for plotting are as follows:

```
library(ggplot2)
library(sp)
library(sf)
library(viridis)
library(patchwork)
```

The package ggplot2 (Wickham 2016) is used for data visualization. The sp (E. J. Pebesma and Bivand 2005) provides tools for handling spatial data. The sf simple features (E. Pebesma 2018) builds upon the strengths of the sp package, introducing efficient approach to handling spatial data. The viridis (Garnier et al. 2023) package provides a collection of color palettes that are color blind friendly. The patchwork (Pedersen 2023) package is used for the combination and arrangement of multiple plots. Figure 1 shows the visualizations of Sri Lanka's administrative borders based on data available in the `ceylon` package. The codes to produce Figure 1 is given below.

```
data(sf_sl_0)
a <- ggplot(sf_sl_0) + geom_sf() + ggtitle("a: Country")
data(province)
b <- ggplot(province) + geom_sf() + ggtitle("b: Province")
data(district)
c <- ggplot(district) + geom_sf() + ggtitle("c: District")
data(sf_sl_3)
d <- ggplot(sf_sl_3) + geom_sf() + ggtitle("d: Divisional Secretariat")
(a|b|c|d)
```

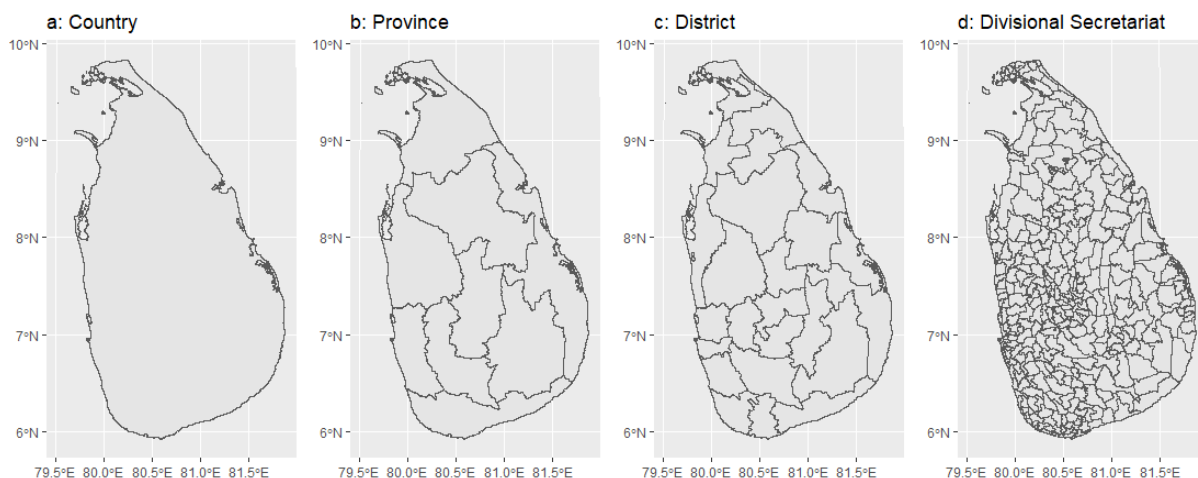


Figure 1: Maps of differnt administrative divisions in Sri Lanka

Point Map: Adding a point to the map

The Global Positioning System (GPS) coordinates of Bandaranaike International Airport, Sri Lanka is Latitude: 7.1753 Longitude: 79.8835. The goal is to plot this point along with the province boundaries. The EPSG:4326 geographic CRS system gives latitude and longitude coordinates to specify a location on the surface of the earth. Hence, before plotting, first longitude and latitudes should be converted into sf object to the same coordinate reference system as the province data set. For this the `sp` (E. J. Pebesma and Bivand 2005) and `sf` (E. Pebesma 2018) packages in R were used. In the following code `st_as_sf` specify the current coordinate reference system for longitude and latitude. The function `st_transform` converts the current CRS to the target CRS. The target CRS is the CRS associated with the province, which is defined as `crs = st_crs(province)` inside the `st_transform` function.

```
airport <- data.frame(lng = 79.8835, lat = 7.1753)
airport.new <- airport |>
  st_as_sf(coords = c("lng", "lat"), crs = 4326) |>
  st_transform(crs = st_crs(province))
```

```
point <- ggplot(province) +
  geom_sf() + geom_sf(data = airport.new, size = 2, col = "darkred") +
  ggtitle("a: Point")
```

Line Map: Plot rivers and streams in Sri Lanka

This section illustrates an example of using the rivers dataset in the package `ceylon`.

```
data("rivers")
line <- ggplot(data = sf_sl_0) +
  geom_sf(fill="#edf8b1", color="#AAAAAA") +
  geom_sf(data=rivers, colour="#253494") +
  labs(title = "b: Line")
```

Polygon Map: Creating a choropleth map

A Choropleth map shows different regions coloured according to the numerical values associated with each individual region.

```
polygon <- ggplot(province) +
  geom_sf(mapping = aes(fill = population)) + scale_fill_viridis() +
  ggtitle("c: Polygon")
```

```
(point|line|polygon)
```

Figure 2 shows the above point, line, and polygon maps created using the data available in the `ceylon` package.

The above examples illustrate the datasets available in the package, which are easily integrated with other companion packages that are widely used in spatial analysis and data visualisation.

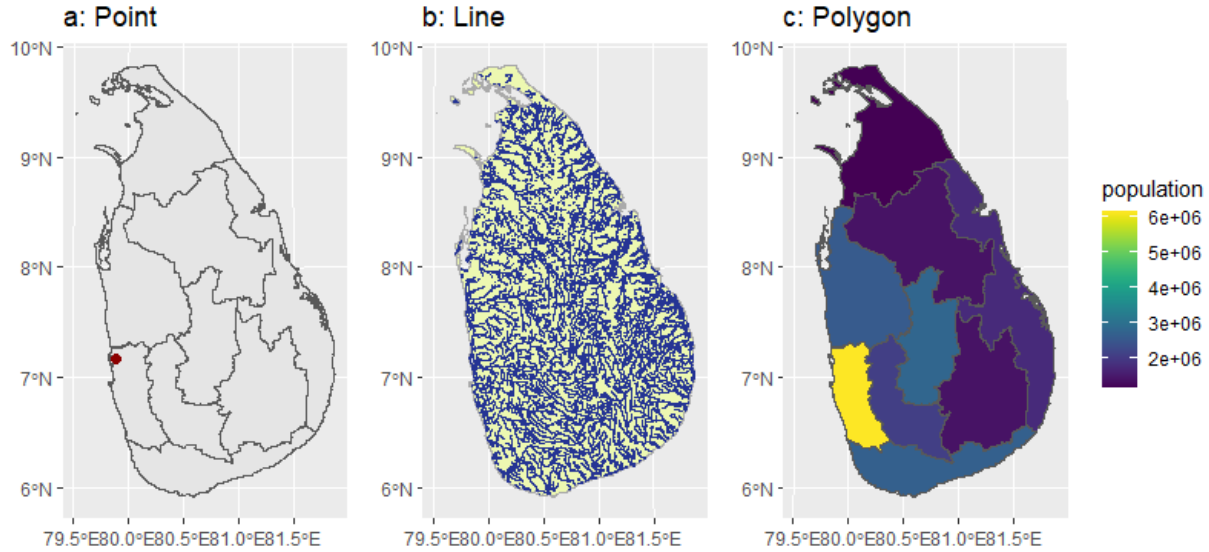


Figure 2: Illustration of point, line and polygon map

Acknowledgements

We extend our sincere gratitude to Stephanie Kobakian for her enlightening talk at the R-Ladies Colombo meetup. This package owes its inspiration to the valuable insights shared during Stephanie’s presentation

References

- Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, et al. 2023. *viridis(Lite) - Colorblind-Friendly Color Maps for r*. <https://doi.org/10.5281/zenodo.4679423>.
- “Humanitarian Data Exchange (HDX).” n.d. Accessed July 23, 2023. <https://data.humdata.org/>.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pebesma, Edzer J., and Roger Bivand. 2005. “Classes and Methods for Spatial Data in R.” *R News* 5 (2): 9–13. <https://CRAN.R-project.org/doc/Rnews/>.
- Pedersen, Thomas Lin. 2023. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Talagala, T. S. 2023. *ceylon: Creating Maps of Sri Lanka Administrative Regions, Rivers and Streams* (version 1.0). Zenodo. <https://doi.org/10.5281/zenodo.10432141>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golem, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.