The Journal of Open Source Software

RCzechia: Spatial Objects of the Czech Republic

Jindra Lacko [1]

1 Prague University of Economics and Business

DOI: 10.21105/joss.05082

Software

- Review
- Repository
- Archive

Editor: Martin Fleischmann

Reviewers:

- @nickbearman
- @paleolimbot

Submitted: 06 December 2022
Published: 02 March 2023

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

`RCzechia` is a R package providing spatial objects and spatial information relevant in the context of the Czech Republic for spatial data analysis and visualization purposes. The package uses `sf` data format to serve the most commonly used administrative areas and natural objects.

As the underlying data is by necessity larger than CRAN package size limits allow, the data is stored externally and a working internet connection is required to use the package.

State of the field

The history of spatial data analysis in `R` is long and respectable Bivand (2021). The first packages focusing specifically on providing spatial data originate from the `S` days Becker & Wilks (1993), with `maps` Deckmyn (2022) being one of the oldest packages in continuous use on CRAN (since 2003). The early packages used pattern of storing spatial data internally, which created a hard limit on volume and level of detail stored.

With the advent of `sp` Pebesma & Bivand (2005) and later `sf` Pebesma (2018) platforms for handling spatial data the universe of packages focused on providing spatial data blossomed. There are packages with global focus, such as `rnaturalearth` South (2017) and regional focus like `giscoR` Hernangómez (2022) oriented at the EU. Number of packages are country specific, such as `tigris` Walker & Rudis (2022) for the US, or `rgugik` Dyba & Nowosad (2021) for Poland. With current near universal and reliable internet access a new pattern has emerged, with spatial data packages accessing cloud stored data files as required (caching them within the limits set by the CRAN repository policy), and distributing only lightweight code.

In the context of Czech Republic there exists `CzechData` package Caha (2021), with somewhat overlapping functionality but available only on GitHub. The CRAN package `czso` Bouchal (2022) interfaces API of the Czech Statistical Office ČSÚ, providing access to statistical data about Czech administrative areas (without the spatial information itself). Package `pragr` Bouchal (2020), available on GitHub, provides geodata about the city of Prague.

Statement of need

No country specific spatial data package has been published on CRAN for the Czech Republic to date, creating a need that could be filled using global or regional packages only to a limited extent.

While there there are open data resources available for researchers, mostly in the format of ESRI Shapefiles, these have a number of practical disadvantages. They have to be located and downloaded individually, and their users in R context face additional hurdles, such as conflicting Coordinate Reference Systems and character encodings. In addition some publicly available datasets are topologically invalid and many are too detailed for use by non GIS specialized audience.

Features

The package provides two distinct sets of spatial objects: administrative areas, and natural objects. In addition API interface wrapping is provided for geocoding and reverse geocoding functions.

Administrative area polygons:

- **republika**: borders of the Czech Republic as a polygon
- **kraje**: 14 regions (NUTS3 areas) of the Czech Republic + Prague as a special case
- **okresy**: 76 districts (LAU1 areas) of the Czech Republic + Prague as a special case
- **orp_polygony** 205 municipalities with extended powers + Prague as a special case
- **obce_polygony**: 6.258 municipalities of the Czech Republic
- **obce_body** the same as obce_polygony, but centroids instead of polygons
- **casti**: primarily 57 city parts of Prague, but also of other cities for which individual boroughs are defined
- **senat_obvody**: 81 Senate districts (upper chamber of Czech Parliament)
- **volebni_okrsky**: 14.761 general election districts
- **zip_codes**: 2.671 ZIP code areas

The country (NUTS1), regions (NUTS3) and districts (LAU1) administrative level objects from `RCzechia` are functionally equivalent to those provided by `giscoR` package Hernangómez (2022) for the Czech Republic. This is expected, as GISCO objects are standardized at the EU level, and the Czech Republic is a EU member state.

Natural objects:

- **reky**: streams and rivers
- **plochy**: stillwaters (lakes and ponds)
- **silnice**: roads (highways, speedways etc.)
- **zeleznice**: railroads
- **chr_uzemi**: protected natural areas
- **lesy**: woodland areas (more than 30 ha in area)
- **KFME_grid**: KFME grid cells according to Niklfeld (1971)
- **vyskopis**: terrain of the Czech republic as a `raster` Hijmans (2022) package object

All objects are implemented as functions returning `sf` class data frames, so commands must be followed by brackets i.e. `RCzechia::republika()`.

For some of the most commonly used objects (*republika*, *kraje*, *okresy*, *reky* and *volebni_okrsky*) an optional low resolution version is also included. To access it, specify the value of `resolution` parameter as `"low"` (default is `"high"`).

Utility functions:

- **geocode**: geocodes an address to coordinates
- **revgeo**: reverse geocodes coordinates to an address

The utility functions interface API of the Czech State Administration of Land Surveying and Cadastre (ČÚZK) and are therefore limited in scope to the area of the Czech Republic.

The package code is thoroughly tested, with 100% test coverage. In addition to testing code the package implements unit tests on integrity of the datasets provided, such as topological validity and internal consistency between different levels of administrative units.

Examples of use

Czech population at the LAU1 level as per the 2011 census, accessed via `czso` package from API of Czech Statistical Office, and mapped using `RCzechia::okresy()` and a `ggplot2` Wickham (2016) call. Note the use of low resolution objects to achieve a more stylized look.

```
library(tidyverse)
library(RCzechia)
library(czso)

src <- czso::czso_get_table("SLDB-VYBER") %>%
   select(uzkod, obyvatel = vse1111) %>%
   mutate(obyvatel = as.numeric(obyvatel))

okresni_data <- RCzechia::okresy("low") %>%
  inner_join(src, by = c("KOD_OKRES" = "uzkod"))

ggplot(data = okresni_data) +
  geom_sf(aes(fill = obyvatel), colour = NA) +
  geom_sf(data = RCzechia::republika("low"), color = "gray30", fill = NA) +
  scale_fill_viridis_c(trans = "log", labels = scales::comma) +
  labs(title = "Czech population",
       fill = "population\n(log scale)") +
  theme_bw() +
  theme(legend.text.align = 1,
        legend.title.align = 0.5)
```

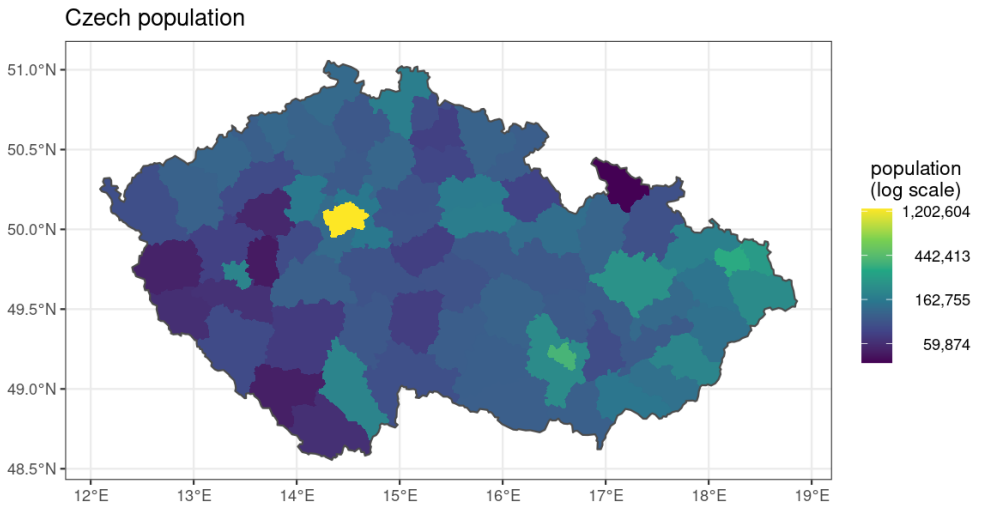

Figure 1: 77 districts of the Czech Republic, with population

Terrain of the Czech Republic, accessed via `RCzechia::vyskopis()` call and displayed using `ggplot2` together with major rivers `RCzechia::reky()` for context.

```
library(tidyverse)
library(RCzechia)

relief <- vyskopis("actual", cropped = TRUE)

ggplot() +
  tidyterra::geom_spatraster(data = relief) +
  geom_sf(data = subset(RCzechia::reky(), Major == T), # major rivers
          color = "steelblue", alpha = .5) +
  scale_fill_gradientn(colors = tidyterra::hypso.colors2(10),
                       labels = scales::label_number(suffix = " m"),
                       limits = c(0, 1550),
                       na.value = NA) +
  labs(title = "Czech Rivers & Their Basins",
       fill = "Altitude") +
  theme_bw() +
  theme(axis.title = element_blank(),
        legend.text.align = 1,
        legend.title.align = 0.5)
```

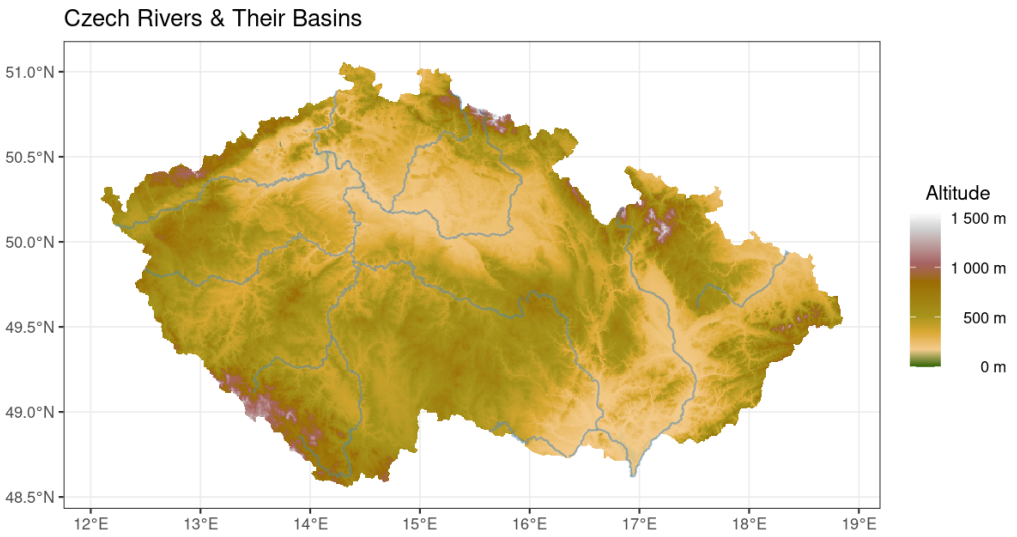

Figure 2: Terrain of the Czech Republic, with major rivers

Examples of `RCzechia` use in current research applications include Korecký et al. (2021) and Brejcha et al. (2021).

References

Becker, R. A., & Wilks, A. R. (1993). *Maps in S*. AT&T Bell Laboratories. https://www.researchgate.net/publication/2684060_Maps_in_S

Bivand, R. S. (2021). Progress in the R ecosystem for representing and handling spatial data. *Journal of Geographical Systems*, *23*(4), 515–546. https://doi.org/10.1007/s10109-020-00336-0

Bouchal, P. (2020). *pragr: Tools for visualising Prague data*. https://github.com/petrbouchal/pragr/

Bouchal, P. (2022). *czso: Use Open Data from the Czech Statistical Office in R*. https://CRAN.R-project.org/package=czso

Brejcha, J., Kodejš, K., Benda, P., Jablonski, D., Holer, T., Chmelař, J., & Moravec, J. (2021). Variability of colour pattern and genetic diversity of Salamandra salamandra (Caudata: Salamandridae) in the Czech Republic. *Journal of Vertebrate Biology*, *70*(2), 21016.1. https://doi.org/10.25225/jvb.21016

Caha, J. (2021). *CzechData: Download various datasets (including spatial data) for the Czech Republic*. https://github.com/JanCaha/CzechData

Deckmyn, O. S. code by R. A. B. &. A. R. W. R. version by R. B. E. by T. P. M. &. A. Deckmyn. (2022). *maps: Draw Geographical Maps*. https://CRAN.R-project.org/package=maps

Dyba, K., & Nowosad, J. (2021). rgugik: Search and Retrieve Spatial Data from the Polish Head Office of Geodesy and Cartography in R. *Journal of Open Source Software*, *6*(59), 2948. https://doi.org/10.21105/joss.02948

Hernangómez, D. (2022). *giscoR: Download map data from GISCO API - Eurostat*. Zenodo. https://doi.org/10.5281/ZENODO.4317946

Hijmans, R. J. (2022). *terra: Spatial Data Analysis*. https://CRAN.R-project.org/package=terra

Korecký, J., Čepl, J., Stejskal, J., Faltinová, Z., Dvořák, J., Lstibůrek, M., & El-Kassaby, Y. (2021). Genetic diversity of Norway spruce ecotypes assessed by GBS-derived SNPs. *Scientific Reports*, *11*, 23119. https://doi.org/10.1038/s41598-021-02545-z

Niklfeld, H. (1971). Bericht uber die Kartierung der Flora Mitteleuropas. *TAXON*, *20*(4), 545–571. https://doi.org/10.2307/1218258

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

Pebesma, E., & Bivand, R. S. (2005). Classes and Methods for Spatial Data: the sp Package. *R News 5*, 21. https://cran.r-project.org/doc/Rnews/

South, A. (2017). *rnaturalearth: World map data from Natural Earth*. https://CRAN.R-project.org/package=rnaturalearth

Walker, K., & Rudis, B. (2022). *tigris: Load census TIGER/line shapefiles*. https://CRAN.R-project.org/package=tigris

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