eurostat: Eurostat Open Data R Tools Leo Lahti, Janne Huovari, Markus Kainu, Przemyslaw Biecek

Governmental institutions have started to release increasing amount of their data resources for the public as open data in the recent years. This is opening novel opportunities for research and citizen science. The eurostat R package provides a suite of tools to access open data from Eurostat, including functions to search, download, and manipulate these data sets in an automated and reproducible manner. The online documentation provides further examples on how to visualize, summarize and interpret these spatio-temporal data sets. The package builds on and extends the preceding SmarterPoland and statfi packages and has been extensively tested by the user community. This package contributes to the growing ecosystem of R packages that provide generic tools for reproducible computational research in social science and humanities.

Efficient tools to access and analyse data collections available in the public domain can greatly benefit reproducible research Gandrud13, Boettiger2015. When the data is available, the complete analytical workflow spanning from raw data to the final publication can be made available. Standardization and automatization of common data analysis tasks via dedicated software packages can help to automate the analysis workflow, thus greatly facilitate reproducibility and code sharing and making data analysis more transparent and efficient.

Here, we introduce the eurostat R package that implements R tools to access open data from Eurostat http://ec.europa.eu/eurostat Eurostat provides a rich collection of demographic and economic data through its open data portal. The portal currently includes ... data sets on ... between years ...

Despite many efforts to this direction, a dedicated R package for eurostat open data has been missing. Our work extends our earlier CRAN packages statfi statfi and smarterpoland smarterpoland. Compared to this earlier work, we have now implemented an expanded set of tools specifically focusing on the eurostat data collection. The eurostat R package hence brings together earlier, independent efforts by the package developers. It has been actively developed by several contributors and based on the community feedback in Github, with its first CRAN release in 2014. We are now reporting the first mature version of the package that has been improved and tested by multiple users, and includes features like cache, handling dates, and using the tidy data principles wickham2014, with the help of the tidyr R package tidyr.

The datamart datamart and the quandl quandl R packages provide generic tools that can be used to access certain versions of eurostat data. In contrast to these generic database packages, our the eurostat package provides functionality that is particularly tailored for this data collection. There is also a development version for R package reurostathttps://github.com/Tungurahua/reurostatut this does not seem to be actively maintained. The package depends on or imports the following external R packages: devtools devtools, dplyr dplyr, knitr knitr, ggplot2 ggplot2, mapproj mapproj, plotrix plotrix, reshape2 reshape2, rmarkdown rmarkdown, stringi stringi, testthat testthat, and tidyr tidyr. The eurostat R package is part of the rOpenGov project Lahti13icml, which provides reproducible research tools for computational social science and digital humanities.

The package provides tools to search and retrieve data from the Eurostat open data portal, to convert identifiers in human-readable formats and to select, modify and visualize the data. In this manuscript, we provide a brief overview of the core functionality in the current CRAN release version (1.2.1). For further examples, see the package vignettehttps://github.com/rOpenGov/eurostat/vignette/eurostat_tutorial.Rmd

To install the CRAN release version, type in R:

install.packages("eurostat")

Search and download commands

The complete table of contents of the database can be downloaded in R with the function get_eurostat_toc()█ [HERE LINK TO EUROSTAT PAGE WHERE THE DATA CAN BE BROWSED ONLINE]. The function search_eurostat() can be used to make a more focused search over the table of contents. To retrieve data sets for 'disposable income', for instance, use:

example library(eurostat) income ¡- $search_eurostat("disposableincome", type = "dataset")$