

rdefra: Interact with the UK AIR Pollution Database from DEFRA

Claudia Vitolo¹, Andrew Russell², and Allan Tucker³

¹Brunel University London

²Brunel University London

³Brunel University London

3 August 2016

Paper DOI: <http://dx.doi.org/10.21105/joss.00051>

Software Repository: https://github.com/kehraProject/r_rdefra

Software Archive: <http://dx.doi.org/10.5281/zenodo.61033>

Summary

Rdefra (Vitolo, Russell, and Tucker 2016) is an R package (R Core Team 2016) to retrieve air pollution data from the Air Information Resource (UK-AIR) of the Department for Environment, Food and Rural Affairs in the United Kingdom. UK-AIR does not provide a public API for programmatic access to data, therefore this package scrapes the HTML pages to get relevant information.

This package follows a logic similar to other packages such as `waterData`(Ryberg and Vecchia 2014) and `rnrf`(Vitolo, Fry, and Buytaert 2015): sites are first identified through a catalogue, data are imported via the station identification number, then data are visualised and/or used in analyses. The metadata related to the monitoring stations are accessible through the function `ukair_catalogue()`, missing stations' coordinates can be obtained using the function `ukair_get_coordinates()`, and time series data related to different pollutants can be obtained using the function `ukair_get_hourly_data()`.

The package is designed to collect data efficiently. It allows to download multiple years of data for a single station with one line of code and, if used with the parallel package (R Core Team 2016), allows the acquisition of data from hundreds of sites in only few minutes.

The figure below shows the 6566 stations with valid coordinates within the UK-AIR (blue circles) database, for 225 of them hourly data is available and their location is shown as red circles.

References

- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ryberg, Karen R., and Aldo V. Vecchia. 2014. *WaterData: An R Package for Retrieval, Analysis, and Anomaly Calculation of Daily Hydrologic Time Series Data*. <https://CRAN.R-project.org/package=waterData>.
- Vitolo, Claudia, Matthew Fry, and Wouter Buytaert. 2015. *Rnrf: UK National River Flow Archive Data from R*. <https://CRAN.R-project.org/package=rnrf>.
- Vitolo, Claudia, Andrew Russell, and Allan Tucker. 2016. "Rdefra: Interact with the UK AIR Pollution Database from DEFRA." doi:10.5281/zenodo.55270.

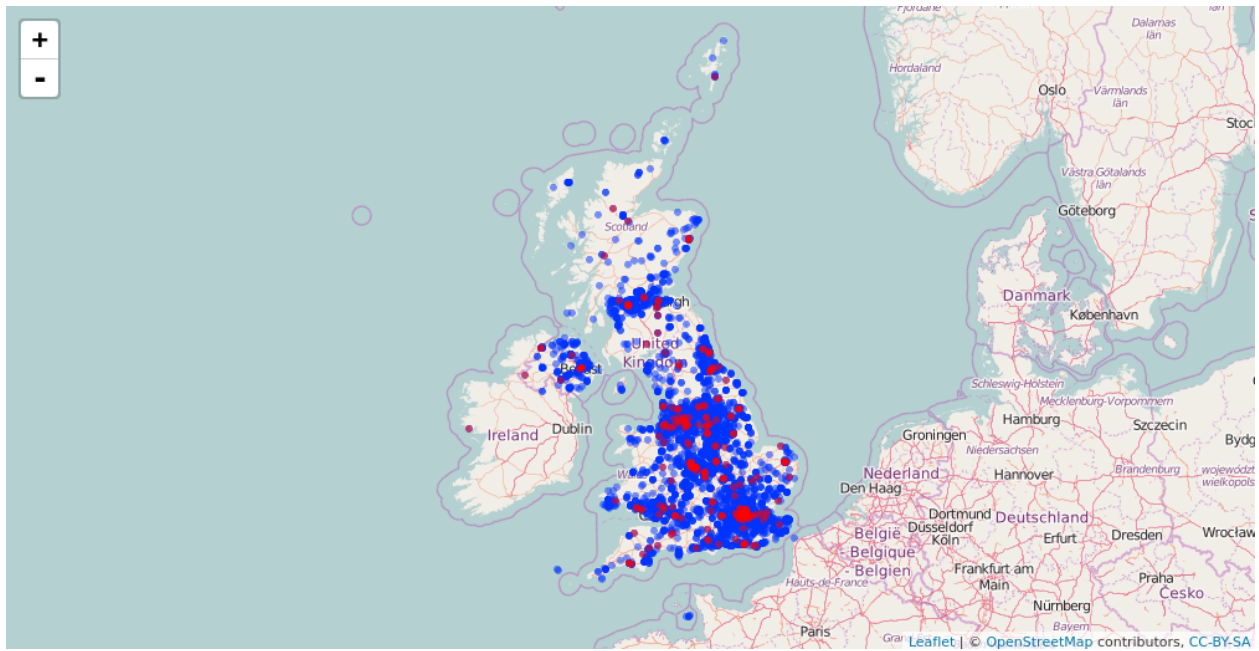


Figure 1: UK-AIR monitoring stations (August 2016)