SOFTWARE TOOL ARTICLE

rspatialdata: a collection of data sources and tutorials on downloading and visualising spatial data using R [version 1; peer review: 2 approved]

Paula Moraga [ID]1, Laurie Baker2

1Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia
2College of the Atlantic, 105 Eden St, Bar Harbor, ME, 04609, USA

Abstract

Spatial and spatio-temporal data are used in a wide range of fields including environmental, health and social disciplines. Several packages in the statistical software R have been recently developed as clients for various databases to meet the growing demands for easily accessible and reliable spatial data. While documentation on how to use many of these packages exist, there is an increasing need for a one stop repository for tutorials on this information. In this paper, we present *rspatialdata* a website that provides a collection of data sources and tutorials on downloading and visualising spatial data using R. The website includes a wide range of datasets including administrative boundaries of countries, Open Street Map data, population, temperature, vegetation, air pollution, and malaria data. The goal of the website is to equip researchers and communities with the tools to engage in spatial data analysis and visualisation so that they can address important local issues, such as estimating air pollution, quantifying disease burdens, and evaluating and monitoring the United Nation's sustainable development goals.

Keywords

Spatial data, open data, visualization, maps, sustainable development goals, R

Corresponding authors: Paula Moraga (paula.moraga@kaust.edu.sa), Laurie Baker (lbaker@coa.edu)

Author roles: Moraga P: Conceptualization, Funding Acquisition, Software, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Baker L**: Writing – Original Draft Preparation, Writing – Review & Editing

Introduction

Spatial data plays a crucial role in a wide range of disciplines, such as environment, health, agriculture, economy and society, and can help governments, companies and citizens improve decision-making. A key example is the use of spatial data by statistical offices worldwide to improve the evaluation and monitoring of the United Nations' Sustainable Development Goals (SDGs) including those related to health, poverty, inequality, climate and the environment.[1]

Spatial data are critical in determining the future of endangered and threatened species,[2] assessing current and future air quality[3] and its effect on population health, and for revealing health inequalities and the early warning of infectious disease outbreaks.[4] For example, mapping and analysis of spatial data are critical in the development of management plans to ensure the efficient use of natural resources such as land and water so that the benefits of these resources can be enjoyed by future generations.[5] Many of these issues do not occur in isolation. Tackling the SDGs requires the integration and combination of data from different sources including social, economic and environmental data. Location often provides the link between these otherwise disparate datasets. High-resolution spatial data is crucial to tailoring management plans to local situations.

The way we monitor change is being rapidly transformed by advances in technology, computing, and data science techniques. Spatial and spatio-temporal data are becoming increasingly common due to advances in both data collection and management. Novel open data sources such as satellite imagery, remote sensing, and Global Positioning System (GPS) data can be collected in large quantities at high spatial and temporal resolutions, at relatively low cost. At the same time, administrative spatial data are becoming increasingly available in open formats. These data are obtained by registries, surveys, and monitoring stations as well as through community-contributed data platforms. Despite a wealth of large and diverse spatial data sources, spatial data may still be hard to find, difficult to use, or not readily accessible. These hurdles limit the re-use of data and their potential impact. These challenges have been recognised for all scientific data, including spatial, and have led to the development of the Findable, Accessible, Interoperable and Reusable (FAIR) guiding principles for scientific data management and stewardship.[6] To maximise their value, data should be FAIR. The first step in (re)using the data is to find them. Therefore, data files should also include descriptive metadata that makes them easily findable for both humans and computers. Once the data are found, users also need to know how data can be accessed, possibly including authentication and authorisation. Data also needs to be interoperable so they can be integrated with other data and interoperate with applications or workflows for analysis, storage and processing. Finally, data should be reusable and to achieve this, they should be well-described so that they can be used and extended in different settings.

R[7] is a powerful language for statistical programming that incorporates a wide range of packages that can be used for data access, manipulation, analysis and visualisation.[8,9] Moreover, R includes several packages that act as clients for various spatial databases and repositories to meet the growing demands for easily accessible and reliable spatial and spatio-temporal data. While documentation and many open source repositories on how to use these packages to access these data sources exist, there is an increasing need for a one stop repository for information about these data sources and tutorials on how to access them using these packages.

Here, we present rspatialdata, a website that presents a collection of reproducible tutorials on how to download, manipulate and visualize a wide range of spatial data including administrative boundaries, population density, climate and health data using the statistical software R. The website makes it easier for individuals to explore, access and use a range of spatial data facilitating the conversion of data into tangible impacts. *rspatialdata* makes these diverse data more Findable and Accessible by grouping instructions together in one place and promoting them to the R community. Interoperability and Reuse are made easier by demonstrating how to read and manipulate the data in a common analysis system with tutorials that promote the reuse of data and analyses.

Methods

Implementation

The tutorials presented in *rspatialdata* have been created using the open-source R Project for Statistical Computing (RRID:SCR_001905)[7] and a number of R packages that allow us to download spatial data corresponding to specific geographic regions and periods of time, as well as to manipulate and visualize the data. Here, we provide a description on how to install the statistical software R and R packages. Then, we show an example on how to download and visualize one of the datasets presented in the website, namely, maximum temperature data. The complete code for all the tutorials can be found at the *rspatialdata* website, and a summary of the datasets and associated R packages included in the website are summarized in Table 1. The code is available from GitHub and is archived with Zenodo.[87]

Table 1. All the datasets included in the *rspatialdata* website and databases and R packages that can be used to retrieve them.

Data	R package	Database
Administrative boundaries	`rgeoboundaries`	geoBoundaries
Population	`wopr`	WorldPop
OpenStreetMap	`osmdata`	OpenStreetMap (OSM)
Elevation	`elevatr`	AWS Terrain Tiles
Temperature	`raster`	WorldClim
Rainfall	`nasapower`	NASA-POWER Project
Humidity	`nasapower`	NASA-POWER Project
Vegetation	`MODIStsp`	Moderate Resolution Imaging Spectroradiometer (MODIS)
Land cover	`MODIStsp`	Moderate Resolution Imaging Spectroradiometer (MODIS)
Air pollution	`openair`	UK Department for Environment Food & Rural Affairs
Demographic and Health Surveys (DHS)	`rdhs`	DHS Program
Malaria	`malariaAtlas`	Malaria Atlas Project (MAP)
Species Occurrence	`spocc`	Global Biodiversity Information Facility (GBIF)

Installation of R and R packages

R[7] is a free, open source, software environment for statistical computing and graphics with many useful packages for importing and manipulating data, statistical modeling, and visualization. R can be downloaded and installed from the Comprehensive R Archive Network (CRAN) (RRID:SCR_003005). R packages can be installed from CRAN with the function `install.packages()` passing the name of the package as first argument in quotes. Then, to use the package, the package needs to be loaded with the function `library()`. For example, we can install and load the visualization package `ggplot2` by typing `install.packages("ggplot2")` and `library(ggplot2)`.

Example of a tutorial: Downloading and visualizing temperature data

The WorldClim (RRID:SCR_010244)[10] database contains global weather and climate data for historical and future conditions at high spatial resolution. These datasets can be easily downloaded with the R package `raster`,[11] which implements several functions for reading, writing, manipulating, analyzing and modeling of spatial data. To use the `raster` package, we first need to install it and load it. Then, to download data, we can use the `getData()` function of the `raster` package by specifying several arguments about the dataset we wish to obtain. For example, to download global maximum temperature, we specify the database name (*e.g.*, `"worldclim"`), the variable we want to download (*e.g.*, `"tmax"`), and the spatial resolution in minutes of a degree as follows.

```
install.packages("raster"); library("raster")
dataset <- getData (name = "worldclim", var = "tmax", res = 10)
```

The downloaded object contains 12 files that correspond to the maximum temperature observed each month. We can manipulate the downloaded object to obtain temperature values for a specific month or average temperature spanning several months, and use other R packages to model and visualize the data.

```
library(ggplot2)

gain(tmax_data) <- 0.1 # Convert temperature to degrees Celsius

# Converting the raster object into a dataframe
tmax_data_may_df <- as.data.frame(tmax_data$tmax5, xy = TRUE, na.rm = TRUE)
rownames(tmax_data_may_df) <- c()
```

```
ggplot(data = tmax_data_may_df, aes(x = x, y = y)) +
geom_raster(aes(fill = tmax5)) +
labs(
title = "Maximum temperature in May",
subtitle = "For the years 1970-2000"
) +
xlab("Longitude") +
ylab("Latitude") +
scale_fill_gradientn(
name = "Temperature (°C)",
colours = c("#0094D1", "#68C1E6", "#FEED99", "#AF3301"),
breaks = c(-20, 0, 20, 40)
)
```

Operation

The software R and RStudio are available for Linux, Mac, and Windows operating systems. It is recommended running these tutorials on a recent version of R (at least R version 4.1.1) and RStudio (at least RStudio version 2021.09.0). R can be downloaded from CRAN, the comprehensive R archive network (https://cran.r-project.org/). CRAN is composed of a set of mirror servers distributed around the world and is used to distribute R and R packages. RStudio is an integrated development environment, or IDE, for R programming. RStudio can be downloaded and installed from http://www.rstudio.com/download. It is recommended updating both R and RStudio at least once a year to keep up to date with the most recent changes.

Use cases

The *rspatialdata* website provides a collection of data sources and tutorials on how to download and visualize spatial data, including administrative boundaries, population, elevation, climatic variables, and health data. These data come from different sources. For example, remote sensing data are acquired by sensors that are not in contact with the target of investigation and can be done, for example, using satellites orbiting the Earth. Remote sensing is used to measure everything from land cover (*e.g.*, water, habitat), environmental phenomena (*e.g.*, elevation, water and sea temperature), to our human footprint (*e.g.*, night light maps). More precise information on a range of environmental and climatic variables such as temperature, rainfall and air pollution can be obtained using monitoring stations placed at specific places that provide ground measurements of these variables during different periods of time. Surveys are also useful to obtain information about health, economy and social characteristics of the population at the local scale. Here, we describe the data sources included in the website, as well as the R packages that allow us to download the data. We also give examples of where these data can be used to solve problems in different disciplines such as health, ecology and the environment.

Administrative boundaries

Administrative boundaries are an essential component for making maps and define the spatial extent needed for electoral, planning and statistical studies. These boundaries, which often guide the spatial scale at which data is collected, offer important context to a wide-range of issues. geoBoundaries[12] is an open license resource database of political administrative boundaries. The R package `rgeoboundaries`[13] is an R client for the geoBoundaries application programming interface (API) that allows us to download administrative boundaries of countries at different administrative levels.

This package has been used as a visualization tool for the study of many different real-world problems, such as mapping coronavirus-19 presence in Vietnam,[14] understanding the impact of Global Environment Facility Projects in Uganda[15] and the influence of travel time to health facilities on stillbirths in Nigeria.[16]

The *rspatialdata* tutorial includes an example of how to retrieve the administrative boundaries of single and multiple countries at different administrative boundary levels. It also covers how to download and visualize these data using the `sf`[17] and `leaflet`[18] packages.

Population

Knowing population sizes and their spatial distributions is crucial for many critical decisions from improving access to health, transportation and energy, to planning and building more resilient and sustainable cities. WorldPop[19] aims to provide an open access archive of spatial demographic datasets with a focus on low and middle income countries (LMICs) to support development, disaster response and health applications.

Population data from WorldPop has been used extensively to map health conditions such as cancer,[20] child growth failure,[21] HIV prevalence,[22] and the burden of cholera[23] in Africa. It has also been used to map local variation in educational attainment in Africa,[24] to evaluate the reduction of tree cover in West African Woodlands[25] and to assess clean air in the context of the SDGs.[26]

The WorldPop Open Population Repository provides access to high-resolution population estimates for individual countries and these data can be obtained with the R package `wopr`.[27] The *rspatialdata* tutorial shows examples on how to use `wopr` to download population data for different countries and administrative levels.

OpenStreetMap (OSM) data

OSM[28] is a collaborative project to create a free editable map of the world. OSM is built by a community of mappers that contribute and maintain global data about roads, trails, cafés, railway stations, and more. OSM data can be used in many ways. For example, as a basemap to put other data into context, for routing or navigation, and for planning or logistics for humanitarian groups, utilities and governments. OSM data have been used in a wide range of applications including flood inundation modeling,[29] air pollution exposure,[30] assessment of socio-economic factors and property prices,[31] and for the study of crime and place.[32]

The package `osmdata`[33] allows us to easily import OSM data in R. The *rspatialdata* tutorial includes an example of how to retrieve OSM data using the `osmdata` by creating a bounding box and a query and how to visualized the data with `ggplot2`, `ggmap`[34] and `leaflet`.[18]

Elevation

Elevation data are important in many different applications. For instance, for environmental problems, elevation data have been used as a tool to study the land cover change over the years, in particular, the evolution of European forest cover.[35] As another example, researchers also have been using elevation data as a complementary source of information in the analysis of species connectivity through genetic structure.[36,37]

For retrieving elevation data from many different regions, one may choose to work with the the `elevatr` package.[38] `elevatr` provides access to elevation data from several web services including the Amazon Web Services Terrain Tiles,[39] the Open Topography Global Datasets API,[40] and the USGS Elevation Point Query Service.[41]

The *rspatialdata* tutorial includes an example of how to retrieve and visualize point elevation data for the USA and raster elevation data from a digital elevation model (DEM) for global elevation data.

Climate data: temperature and precipitation

WorldClim[10] is a database that provides high spatial resolution global weather and climate data for historical and future conditions. For example, it provides monthly climate data for minimum, mean, and maximum temperature, precipitation, solar radiation, wind speed, water vapor pressure, and for total precipitation.

These data may be applicable in many different areas. For environmental problems, it has been used for the study of the global tree restoration potential,[42] the understanding of temperature profile in forest regions,[43] and the monitoring of drought in South Asia.[44] In ecology, to understand geographic distribution of sloths in Costa Rica.[2] In health and disease-control related problems, these data have been used, for example, in the study of the levels of arsenic in groundwater,[45] the prediction of lymphatic filariasis prevalence in sub-Saharan Africa,[46] and the loss of biodiversity on Earth due to the amphibian chytridiomycosis panzootic disease.[47]

The package `raster`[11] allows us to easily download the WorldClim data as well as to manipulate and analyze spatial datasets. The rspatialdata tutorial includes an example of how to retrieve maximum temperature data from the WorldClim database and visualize the monthly maximum and mean monthly temperature and other bioclimatic variables over time using `ggplot2` and the `sf` package.[17]

Rainfall and humidity

The NASA Prediction Of Worldwide Energy Resources (POWER) Project[48] provides meteorology, surface solar energy and climatology data for support of renewable energy, building energy efficiency and agricultural needs. Data retrieved from the NASA POWER Project have been used in a few different applications. For example, POWER data have been used in the study of the potential utilization of wind electric pumping systems for water distribution in Cameroon,[49] in the analysis of photovoltaic systems usage in China[50] and in the study of *Dunaliella salina* (a type of green micro-algae) cultivation.[51]

nasapower[52] aims to make it quick and easy to automate downloading NASA-POWER data in R. In *rspatialdata*, we show how to use this package to download rainfall and humidity.

Vegetation and land cover

Vegetation data are used in a wide variety of applications ranging from environmental applications, such as the rice crop monitoring in Europe,[53] to health and disease-control applications, such as malaria transmission dynamics in an indigenous province in Panama.[54]

Vegetation data are captured using Moderate Resolution Imaging Spectroradiometer (MODIS), an instrument onboard the Terra and Aqua NASA scientific research satellites. MODIS captures data in 36 spectral bands in three spatial resolutions across the surface of the earth. Data products derived from these observations include features of the atmosphere, land, cryosphere, and ocean, made available at different frequencies and spatial resolutions. Each data product contains multiple product layers, including original MODIS layers, quality layers and spectral indexes, produced at different intervals and at different spatial resolutions. User guides on each of the product areas are available, which provide in-depth explanations on them.

The *rspatialdata* tutorial shows how to use the R package MODIStsp,[55] which acts as a client for downloading time series and raster images derived from MODIS Land Product data. Specifically, it shows how to download MODIS Vegetation Index Products (NDVI and EVI)[56] and the MODIS Land Cover Products.[57]

Air pollution

Air pollution data can be of interest for many different agents, from the government to the general population. In this sense, many different studies have been conducted regarding how the UK and other countries have been suffering from different types of pollutants—for instance, on how wood-burning has impacted the PM_{10} levels in London,[58] or how the level of air pollution has a direct impact on the population's health,[3] or even how people from different socioeconomic groups may be exposed to different levels of air pollution depending on their commute in London.[59]

UK Air is a UK air quality database provided by the Department for Environment Food & Rural Affairs.[60] The database provides daily information about the level of pollution for different pollutants (*e.g.*, ozone, carbon monoxide, $PM_{2.5}$) across the United Kingdom and its territories. Although there are many different ways to retrieve data from this database, one convenient option is using the openair[61] R package.

The openair package provides a set of functions to import and work with these datasets, which are documented in the openair's manual.[62] The *rspatialdata* tutorial includes an example of how to retrieve and visualize data from a specific monitoring network named Automatic Urban and Rural Network (AURN).

Demographic and Health Surveys (DHS)

The Demographic and Health Surveys (DHS) Program[63] collects, analyzes, and disseminates country-wide subnational level data on population, health, nutrition and HIV. The objective of the DHS Program is to improve and institutionalize the collection and use of data by developing countries for program monitoring and evaluation and for policy making. The R package rdhs[64] provides a wrapper to the DHS program API, and can be used to identify particular datasets and download them in R *via* the DHS API. Examples of issues that have been investigated using DHS data include household smoke-exposure risks associated with cooking fuels and cooking places in Tanzania,[65] determinants of unmet need for family planning and implications for women's health in Gambia & Mozambique,[66] and household access to improved drinking water sources and toilet facilities in Ethiopia.[67]

The *rspatialdata* tutorial includes different examples of options on how to retrieve datasets and DHS surveys for an analysis through the DHS API and DHS website from R. And how to search for a specific DHS survey using tag words demonstrating how to extract surveys on Malaria in Rwanda and Tanzania as a case study.

Malaria

The Malaria Atlas Project (MAP)[68] aims to better understand the global landscape of malaria risk, how this is changing, and the impact of malaria interventions to support malaria intervention and eradication efforts. As part of its work, MAP assembles an extensive collection of malaria data, including parasite rate data (*Plasmodium falciparum* and *Plasmodium vivax*), vector occurrence, and satellite images capturing conditions that influence malaria transmission. malariaAtlas[69] is an R package to open-access malaria data hosted by MAP and can be used to download all publicly available parasite rate survey points, mosquito occurrence points and raster surfaces from the MAP servers as well as utility functions for plotting the downloaded data. Data provided by malariaAtlas can be used to explore the spatial

and spatio-temporal patterns of malaria risk as well as to feed into spatial models of the risk of malaria. Several studies have used MAP data for different purposes, including mapping the global endemicity and clinical burden malaria,[70] understand the associated patterns of insecticide resistance in field populations of malaria vectors across Africa,[71] and assess the population coverage of artemisinin-based combination treatment and *Plasmodium falciparum* infection in Africa.[72]

The *rspatialdata* tutorial includes examples of how to retrieve and visualize malaria data from the `malariaAtlas` package including parasite rate (PR) survey data, vector occurrence data, and rasters of modelled malaria research outputs.

Species occurrence

The information of observed species play an import role in ecological studies, which motivates the existence of different repositories containing these type of data. Examples include GBIF - Global Biodiversity Information Facility (RRID: SCR_005904),[73] Biodiversity Information Serving Our Nation (BISON),[74] eBird,[75] and VertNet.[76] Most of these repositories allow researchers to retrieve data using different methods. In R, the aforementioned platforms can be accessed through the `rgbif`,[77] `rbison`,[78] `rebird`,[79] and `rvertnet`[80] packages, respectively. However, in order to integrate all these datasets and interact with them using just one tool, one could choose to work with the `spocc` package.[81] As an example, and aiming to model sloths occurrence in Costa Rica, `spocc` was used to retrieve relevant data from GBIF.[82] Other case studies may include modeling migratory movements of birds[83] or estimating population size based on species occurrence.[84]

The *rspatialdata* tutorial includes an example of how to retrieve and visualize species occurrence data by creating a query for a species latin name using the `spocc` package.

Discussion

Open and reliable data are crucial for solving global challenges and monitoring the UN Sustainable Development Goals by 2030, including those for improving health, reducing inequalities, and protecting the environment. Accessible spatial data in particular are key to understanding diverse questions ranging from disease spread to climatic trends and necessary for evaluating the impact of interventions and policy decisions.

In this paper, we present *rspatialdata*, a website containing a collection of data sources and tutorials on downloading and visualising spatial data using the statistical software R. The website represents an important step towards helping users find, access and visualize spatial data. As a one-stop repository for tutorials on accessing spatial data, we aim to provide an overview for users on what spatial data is available and how it can be accessed from R. We use motivating examples in the tutorials to illustrate how a variety of spatial data can be used to inform evidence-based decision-making in a wide range of fields. The *rspatialdata* website is a useful resource for individuals working with problems that require spatial data analysis and visualisation, such as estimating air pollution, quantifying disease burdens, predicting species occurrences, and evaluating and monitoring the UN Sustainable Development Goals.

An ongoing challenge in many disciplines that use spatial data is a lack of data in some locations and periods of time, as well as a lack of disaggregated data corresponding to age groups, genders and other factors. Spatial data are often aggregated at the scale of administrative units rather than locally relevant scales. These limitations make it difficult to compare processes over time and to evaluate outcomes for different population groups. While modeling techniques can be used to fill these gaps,[85,86] it is important to continue supporting countries to generate and access data that will help inform better decision-making globally.

We have chosen to write tutorials for spatial datasets that are important for decision-making in a wide range of fields such as health, climate, environment and ecology. While there may be different packages that do the same as the packages included in the website, *rspatialdata* tries to present the packages that are easiest to install and use, and includes other additional packages in the reference sections so users can explore additional functionalities and examples these packages provide. The website will be updated by including noteworthy packages to retrieve spatial data as they are discovered, and tutorials of existing packages will be updated if the code to use them changes or there are new notable functions to include. Also, in order to encourage the community to contribute, the website provides guidelines for contribution. The *rspatialdata* website is not comprehensive and it does not contain all available datasets. Nevertheless, it can provide a useful resource to get users started and a stimulus and location for others to contribute.

We expect the quantity and variety of spatial data provided by novel data streams such as satellite imagery, remote sensing, and GPS tracking to only increase in the future. The *rspatialdata* website will be regularly updated to meet the

Table 2. Databases included in the *rspatialdata* website.

Data	Database
Administrative boundaries	geoBoundaries
Population	WorldPop
OpenStreetMap	OpenStreetMap (OSM)
Elevation	AWS Terrain Tiles
Temperature	WorldClim
Rainfall	NASA-POWER Project
Humidity	NASA-POWER Project
Vegetation	Moderate Resolution Imaging Spectroradiometer (MODIS)
Land cover	Moderate Resolution Imaging Spectroradiometer (MODIS)
Air pollution	UK Department for Environment Food & Rural Affairs
Demographic and Health Surveys (DHS)	DHS Program
Malaria	Malaria Atlas Project (MAP)
Species Occurrence	Global Biodiversity Information Facility (GBIF)

growing demands to access spatial data by the R community and to include new R packages and data sources as they are developed and released. By promoting the reuse and sharing of spatial data and spatial analyses, the *rspatialdata* website contributes to community-building and sharing of best practices on working with spatial data.

Data availability
Underlying data
Table 2 contains the databases included in the *rspatialdata* website.

Software availability
Software available from: https://rspatialdata.github.io/.

Source code available from: https://github.com/rspatialdata/rspatialdata.github.io.

Archived source code at time of publication: https://doi.org/10.5281/zenodo.6779351.[87]

License: MIT

References

1. Assembly, General: **Resolution adopted by the general assembly on 19 september 2016. Technical report, A/RES/71/1, 3 October 2016 (The New York Declaration).** 2015.

2. Moraga P: **Species Distribution Modeling using Spatial Point Processes: a Case Study of Sloth Occurrence in Costa Rica.** *R J.* 2021; **12**(2): 293–310.
Publisher Full Text

3. Heal MR, Kumar P, Harrison RM: **Particles, air quality, policy and health.** *Chem. Soc. Rev.* 2012; **41**(19): 6606–6630.
Publisher Full Text

4. Moraga P, Dorigatti I, Kamvar ZN, *et al.*: **epiflows: an R package for risk assessment of travel-related spread of disease.** *F1000Res.* 2018; **7**: 1374.
Publisher Full Text

5. Liping C, Yujun S, Saeed S: **Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques - A case study of a hilly area, Jiangle, China.** *PLoS One.* 2018; **13**(7): e0200493.
PubMed Abstract | **Publisher Full Text**

6. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Scientific Data.* March 2016; **3**(1): 160018. 2052-4463. Number: 1 Publisher: Nature Publishing Group.
PubMed Abstract | **Publisher Full Text** | **Reference Source**

7. R Core Team: *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2018.
Reference Source

8. Moraga P: *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny.* CRC Press; 2019.

9. Moraga P: **SpatialEpiApp: A Shiny Web Application for the analysis of Spatial and Spatio-Temporal Disease Data.** *Spatial and Spatio-temporal Epidemiology.* 2017; **23**: 47–57.
PubMed Abstract | **Publisher Full Text**

10. WorldClim: **Global climate and weather data.**
Reference Source

11. Robert J: **Hijmans.** *raster: Geographic Data Analysis and Modeling.* 2020. R package version 3.4-5.
Reference Source

12. Runfola D, Anderson A, Baier H, *et al.*: **geoboundaries: A global database of political administrative boundaries.** *PLoS One.* 04 2020; **15**(4): 1–9.
PubMed Abstract | **Publisher Full Text**

13. Dicko A: **rgeoboundaries: A Client to geoBoundaries, A Political Administrative Boundaries Dataset.** 2020. R package version 0.0.0.9000.
Reference Source

14. Huong NQ, Nga NTT, Van Long N, *et al.*: **Coronavirus testing indicates transmission risk increases along wildlife supply chains for human consumption in viet nam, 2013-2014.** *PLoS One.* 2020; **15**(8): e0237129.
PubMed Abstract | **Publisher Full Text**

15. Runfola D, Batra G, Anand A, *et al.*: **Exploring the socioeconomic co-benefits of global environment facility projects in uganda using a quasi-experimental geospatial interpolation (QGI) approach.** *Sustainability.* 2020; **12**(8): 3225.
Publisher Full Text

16. Wariri O, Onuwabuchi E, Alhassan JAK, *et al.*: **The influence of travel time to health facilities on stillbirths: A geospatial case-control analysis of facility-based data in gombe, nigeria.** *PLoS One.* 2021; **16**(1): e0245297.
PubMed Abstract | **Publisher Full Text**

17. Pebesma E: **Simple Features for R: Standardized Support for Spatial Vector Data.** *R J.* 2018; **10**(1): 439–446.
Publisher Full Text

18. Cheng J, Karambelkar B, Xie Y; *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library.* 2022. R package version 2.1.0.
Reference Source

19. *WorldPop (School of Geography and Environmental Science, University of Southampton).* 2021.
Reference Source

20. Moraga P: **Small area disease risk estimation and visualization using r.** *R J.* 2018; **10**: 495–506.
Publisher Full Text

21. Osgood-Zimmerman A, Millear AI, Stubbs RW, *et al.*: **Mapping child growth failure in africa between 2000 and 2015.** *Nature.* 2018; **555**(7694): 41–47.
PubMed Abstract | **Publisher Full Text**

22. Dwyer-Lindgren L, Cork MA, Sligar A, *et al.*: **Mapping hiv prevalence in sub-saharan africa between 2000 and 2017.** *Nature.* 2019; **570**(7760): 189–193.
PubMed Abstract | **Publisher Full Text**

23. Lessler J, Moore SM, Luquero FJ, *et al.*: **Mapping the burden of cholera in sub-saharan africa and implications for control: an analysis of data across geographical scales.** *Lancet.* 2018; **391**(10133): 1908–1915.
PubMed Abstract | **Publisher Full Text**

24. Graetz N, Friedman J, Osgood-Zimmerman A, *et al.*: **Mapping local variation in educational attainment across africa.** *Nature.* 2018; **555**(7694): 48–53.
PubMed Abstract | **Publisher Full Text**

25. Brandt M, Rasmussen K, Hiernaux P, *et al.*: **Reduction of tree cover in west african woodlands and promotion in semi-arid farmlands.** *Nat. Geosci.* 2018; **11**(5): 328–333.
PubMed Abstract | **Publisher Full Text**

26. Rafaj P, Kiesewetter G, Gül T, *et al.*: **Outlook for clean air in the context of sustainable development goals.** *Glob. Environ. Chang.* 2018; **53**: 1–11.
Publisher Full Text

27. Leasure DR, Bondarenko M, Darin E, *et al.*: **wopr: An R package to query the WorldPop Open Population Repository, version 0.4.5.** 2020.
Reference Source

28. OpenStreetMap contributors: **Planet dump.** 2017.
Reference Source | **Reference Source** .

29. Hawker L, Rougier J, Neal J, *et al.*: **Implications of simulating global digital elevation models for flood inundation studies.** *Water Resour. Res.* 2018; **54**(10): 7910–7928.
Publisher Full Text

30. Ramacher MOP, Karl M: **Integrating modes of transport in a dynamic modelling approach to evaluate population exposure to ambient no2 and pm2. 5 pollution in urban areas.** *Int. J. Environ. Res. Public Health.* 2020; **17**(6): 2099.
PubMed Abstract | **Publisher Full Text**

31. Dupré D: **Urban and socio-economic correlates of property prices in dublin's area.** *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA).* pages 556–562. IEEE. 2020.

32. Langton S, Solymosi R: **Open street map for crime and place.** 2020.

33. Padgham M, Rudis B, Lovelace R, *et al.*: **osmdata.** *J. Open Source Softw.* jun 2017; **2**(14).
Publisher Full Text

34. Kahle D, Wickham H: **ggmap: Spatial visualization with ggplot2.** *R J.* 2013; **5**(1): 144–161.
Publisher Full Text | **Reference Source**

35. Zanon M, Davis BAS, Marquer L, *et al.*: **European forest cover during the past 12,000 years: a palynological reconstruction based on modern analogs and remote sensing.** *Front. Plant Sci.* 2018; **9**: 253.
PubMed Abstract | **Publisher Full Text**

36. O'Connell KA, Mulder KP, Maldonado J, *et al.*: **Sampling related individuals within ponds biases estimates of population structure in a pond-breeding amphibian.** *Ecol. Evol.* 2019; **9**(6): 3620–3636.
PubMed Abstract | **Publisher Full Text**

37. Mulder KP, Cortes-Rodriguez N, Campbell EH, *et al.*: **North-facing slopes and elevation shape asymmetric genetic structure in the range-restricted salamander plethodon shenandoah.** *Ecol. Evol.* 2019; **9**(9): 5094–5105.
PubMed Abstract | **Publisher Full Text**

38. Hollister J, Shah T, Robitaille AL, *et al.*: **elevatr: Access Elevation Data from Various APIs.** 2020. R package version 0.3.1.
Publisher Full Text | **Reference Source**

39. Registry of Open Data on AWS: **Terrain Tiles.**
Reference Source

40. Open Topography: **Open Topography API.**
Reference Source

41. United States Geological Survey: **The National Map.** 2017.
Reference Source

42. Bastin J-F, Finegold Y, Garcia C, *et al.*: **The global tree restoration potential.** *Science.* 2019; **365**(6448): 76–79.
Publisher Full Text

43. De Frenne P, Zellweger F, Rodriguez-Sanchez F, *et al.*: **Global buffering of temperatures under forest canopies.** *Nat. Ecol. Evol.* 2019; **3**(5): 744–749.
PubMed Abstract | **Publisher Full Text**

44. Aadhar S, Mishra V: **High-resolution near real-time drought monitoring in south asia.** *Scientific Data.* 2017; **4**(1): 1–14.
Publisher Full Text

45. Podgorski J, Berg M: **Global threat of arsenic in groundwater.** *Science.* 2020; **368**(6493): 845–850.
PubMed Abstract | **Publisher Full Text**

46. Moraga P, Cano J, Baggaley RF, *et al.*: **Modelling the distribution and transmission intensity of lymphatic filariasis in sub-saharan africa prior to scaling up interventions: integrated use of geostatistical and mathematical modelling.** *Parasit. Vectors.* 2015; **8**(1): 560.
PubMed Abstract | **Publisher Full Text**

47. Scheele BC, Pasmans F, Skerratt LF, *et al.*: **Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity.** *Science.* 2019; **363**(6434): 1459–1463.
PubMed Abstract | **Publisher Full Text**

48. NASA: **The POWER Project.**
Reference Source

49. Kidmo DK, Bogno B, Deli K, *et al.*: **Economic assessment of wecs for water pumping systems in the north region of cameroon.** *Renew. Energy Environ. Sustain.* 2021; **6**: 6.
Publisher Full Text

50. Liang J, Gao X: **Assessing the regional grid-parity potential of utility-scale photovoltaic in china.** *IOP Conference Series: Earth and Environmental Science.* IOP Publishing; 2020; volume **512**: page 012022.

51. Borovkov AB, Gudvilovich IN, Avsiyan AL: **Scale-up of dunaliella salina cultivation: from strain selection to open ponds.** *J. Appl. Phycol.* 2020; **32**(3): 1545–1558.
Publisher Full Text

52. Sparks A: *nasapower: NASA-POWER Data from R.* 2020. R package version 3.0.1.
Reference Source

53. Busetto L, Casteleyn S, Granell C, *et al.*: **Downstream services for rice crop monitoring in europe: From regional to local scale.** *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* 2017; **10**(12): 5423–5441.
Publisher Full Text

54. Hurtado LA, Calzada JE, Rigg CA, *et al.*: **Climatic fluctuations and malaria transmission dynamics, prior to elimination, in guna yala, república de panamá.** *Malar. J.* 2018; **17**(1): 1–12.
Publisher Full Text

55. Busetto L, Ranghetti L: **Modistsp: an r package for preprocessing of modis land products time series.** *Comput. Geosci.* 2016; **97**:

40–48. 0098-3004.
Publisher Full Text | **Reference Source**

56. Didan K: **MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC.** 2015.
Publisher Full Text

57. Friedl M, Sulla-Menashe D: **MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.**
Publisher Full Text

58. Fuller GW, Tremper AH, Baker TD: **Karl Espen Yttri, and David Butterfield. Contribution of wood burning to pm10 in london.** *Atmos. Environ.* 2014; **87**: 87–94.
Publisher Full Text

59. Rivas I, Kumar P, Hagen-Zanker A: **Exposure to air pollutants during commuting in london: are there inequalities among different socio-economic groups?** *Environ. Int.* 2017; **101**: 143–157.
PubMed Abstract | **Publisher Full Text**

60. UK-Air: **UK Department for Environment, Food & Rural Affairs.**
Reference Source .

61. Carslaw DC, Ropkins K: **openair—An R package for air quality data analysis.** *Environ. Model Softw.* 2012; **27-28** (0): 52–61.
Publisher Full Text

62. Carslaw DC: *The openair book—Tools for air quality data analysis.* 2020.
Reference Source

63. ICF: *The DHS Program Spatial Data Repository. Funded by USAID.*
Reference Source

64. Watson OJ, FitzJohn R, Eaton JW: **rdhs: an r package to interact with the demographic and health surveys (dhs) program datasets.** *Wellcome Open Res.* 2019; **4**: 103.
Publisher Full Text | **Reference Source**

65. Ahamad MG, Tanin F, Shrestha N: **Household smoke-exposure risks associated with cooking fuels and cooking places in tanzania: A cross-sectional analysis of demographic and health survey data.** *Int. J. Environ. Res. Public Health.* 2021; **18**(5): 2534.
PubMed Abstract | **Publisher Full Text**

66. Yaya S, Idriss-Wheeler D, Uthman OA, *et al.*: **Determinants of unmet need for family planning in gambia & mozambique: implications for women's health.** *BMC Womens Health.* 2021; **21**: 123.
PubMed Abstract | **Publisher Full Text**

67. Andualem Z, Dagne H, Azene ZN, *et al.*: **Households access to improved drinking water sources and toilet facilities in ethiopia: a multilevel analysis based on 2016 ethiopian demographic and health survey.** *BMJ Open.* 2021; **11**: e042071.
PubMed Abstract | **Publisher Full Text**

68. Hay SI, Snow RW: **The malaria atlas project: Developing global maps of malaria risk.** *PLoS Med.* 12 2006; **3**(12): e473–e475.
PubMed Abstract | **Publisher Full Text**

69. Pfeffer D, Lucas T, May D, *et al.*: **malariaatlas: an r interface to global malariometric data hosted by the malaria atlas project.** *Malar. J.* 2018; **17**(1): 352.
PubMed Abstract | **Publisher Full Text**

70. Battle KE, Lucas TCD, Nguyen M, *et al.*: **Mapping the global endemicity and clinical burden of plasmodium vivax, 2000–17: a spatial and temporal modelling study.** *Lancet.* 2019; **394** (10195): 332–343. 0140–6736.
PubMed Abstract | **Publisher Full Text**

71. Hancock PA, Wiebe A, Gleave KA, *et al.*: **Associated patterns of insecticide resistance in field populations of malaria vectors across africa.** *Proc. Natl. Acad. Sci.* 2018; **115** (23): 5938–5943. 0027-8424.
PubMed Abstract | **Publisher Full Text**

72. Bennett A, Bisanzio D, Yukich JO, *et al.*: **Population coverage of artemisinin-based combination treatment in children younger than 5 years with fever and plasmodium falciparum infection in africa, 2003-2015: a modelling study using data from national surveys.** *Lancet Glob. Health.* 2017; **5** (4): e418–e427. 2214-109X.
PubMed Abstract | **Publisher Full Text**

73. Global Biodiversity Information Facility: **Global Biodiversity Information Facility (GBIF).** 2021.
Reference Source

74. Biodiversity Information Serving Our Nation: **Biodiversity Information Serving Our Nation (BISON).** 2021.
Reference Source

75. Cornell Lab of Ornithology: **eBird.** 2021.
Reference Source

76. VertNet: **VertNet.** 2021.
Reference Source

77. Chamberlain S, Oldoni D, Barve V, *et al.*: **rgbif: Interface to the Global'Biodiversity' Information Facility API.** 2021.
Reference Source

78. Chamberlain S: **rbison: Interface to the 'USGS"BISON' API.** 2020.
Reference Source

79. Maia R, Chamberlain S, Teucher A, *et al.*: **rebird: R Client for the eBird Database of Bird Observations.** 2021.
Reference Source

80. Chamberlain S, Ray C, Barve V: **rvertnet: Search 'Vertnet', a 'Database' of Vertebrate Specimen Records.** 2021.
Reference Source

81. Chamberlain S, Ram K, Hart T: **spocc: Interface to Species Occurrence Data Sources.** 2021.
Reference Source

82. Moraga P: **Species distribution modeling using spatial point processes: a case study of sloth occurrence in costa rica.** *R J.* 2020; **12**(2): 311–320.

83. Walker J, Taylor P: **Using ebird data to model population change of migratory bird species.** *Avian Conserv. Ecol.* 2017; **12**(1).
Publisher Full Text

84. Dorazio RM, Royle JA: **Estimating size and composition of biological communities by modeling the occurrence of species.** *J. Am. Stat. Assoc.* 2005; **100**(470): 389–398.
Publisher Full Text

85. Moraga P, Cramb SM, Mengersen KL, *et al.*: **A geostatistical model for combined analysis of point-level and area-level data using inla and spde.** *Spatial Statistics.* 2017; **21**: 27–41.
Publisher Full Text

86. Moraga P, Ozonoff A: **Model-based imputation of missing data from the 122 Cities Mortality Reporting System (122 CMRS).** *Stoch. Env. Res. Risk A.* 2015; **29**: 1499–1507.
Publisher Full Text

87. Moraga P: **First release (v1.0.0) rspatialdata/rspatialdata.github. io. [Software].** 2022.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 09 August 2022

https://doi.org/10.5256/f1000research.134795.r144387

✔ **Natalia da Silva** iD

Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay

rspatialdata: a collection of data sources and tutors on downloading and visualizing spatial data using R.

This paper presents a website that provides a collection of data sources and tutorials on downloading and visualizing spatial data using R. There are several R packages that simplify the access and reliability of spatial data but rspatialdata can help researchers to find and use spatial data in a simple way, tutorials are focused on read and manipulate data in a common analysis system promoting the data reuse and analyses.

General comment:

Since you are presenting a webpage I think it will be good to include a general description of the webpage structure, describing each tabs, and at least an image with the Home tab describing the user interaction. In Methods before Table 1. Also, you should mention which tools you have used to design the webpage.

It will be good to include a complete use case with all the visualizations as you have on the webpage, maybe on page 4 you can extend the temperature data example and show and comment on the figures, not just the R code.

Minor comments:
1. "R is a powerful language for statistical programming that incorporates a wide range of packages that can be used for data access, manipulation, analysis, and visualization." This is a general statement not focused on spatial data and your references are specific and auto references to your work, you can include some general references or be specific to spatial data.

2. Page 3. It will be useful to include the webpage URL at least once https://rspatialdata.github.io for printed version. The same for the GitHub repo (

https://github.com/rspatialdata/rspatialdata.github.io).

3. Page 4 Table 1, second column contains the R packages, and since this is the first time you mentioned them in the paper I think you should include citation there. Not sure if there is any restrictions on including references in a table in this publishing platform.

4. Page 4 first paragraph, you should cite ggplot2.

5. Remove Table 2, it has the same info as Table 1.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I'm a Statistician interested in statistical computing, data visualization among others.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 29 July 2022

https://doi.org/10.5256/f1000research.134795.r144385

✔   **Emmanuel Olamijuwon** 🆔

University of Southampton, Southhampton, UK

This paper describes a web platform (rspatialdata) that makes diverse population, health, climate and environmental data more Findable and Accessible. The website also provides instructions for accessing, exploring, and reusing (visualising) the datasets. As the authors note, high-resolution spatial data is crucial to tailoring management plans and service delivery to local situations. By promoting the reuse and sharing of spatial data and spatial analyses, the rspatialdata website contributes to community-building and sharing of best practices for working with spatial data.

I agree with the authors that there's an increasing need for a central repository for information about spatial data sources and tutorials on their use. The tutorials are purposefully designed and well written in such a way that they will be easy to understand by anyone with basic r-programming experience. The rspatialdata website is also user-friendly, making it easy for anyone with an internet-enabled device to access them. In recognition of the ever-increasing variety of spatial data provided by novel data streams such as satellite imagery, remote sensing, and GPS tracking, the authors have also included a dedicated section for inviting community contributions, which is commendable.

I have included a few minor comments below with the hopes that they would further strengthen the work.
- Table 1: I suggest changing Demographic and health survey to demographics and health (data focus).

- Table 2 is technically a repetition of Table 1.

- Considering the article's peculiar focus on FAIR principles. It will be important to include a tutorial that demonstrates the interoperability of the datasets. That is, linking demographics and health data to Administrative boundaries or Humidity and/or Population data. I believe this would be a more substantial contribution of this article and website.

- Please include some spatial tutorials for the DHS module/page (https://rspatialdata.github.io/dhs-data.html).

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the

findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Demography and Social Statistics, Global Health; Africa

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research