

rAvis: An R-Package for Downloading Information Stored in Proyecto AVIS, a Citizen Science Bird Project

Sara Varela^{1,2*}, Javier González-Hernández³, Eduardo Casabella⁴, Rafael Barrientos⁵

1 Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic, **2** Departamento de Ecología, Instituto de Ciencias Biológicas, Universidade Federal de Goiás, CxP 131, Goiânia, Goiás, Brasil, **3** Toledo, Spain, **4**ACCONTE Group, Actividades de Consultoría y Telecomunicaciones, Tres Cantos, Spain, **5** Departamento de Ciencias Ambientales, Facultad de Medio Ambiente, Universidad de Castilla-La Mancha, Toledo, Spain

Abstract

Citizen science projects store an enormous amount of information about species distribution, diversity and characteristics. Researchers are now beginning to make use of this rich collection of data. However, access to these databases is not always straightforward. Apart from the largest and international projects, citizen science repositories often lack specific Application Programming Interfaces (APIs) to connect them to the scientific environments. Thus, it is necessary to develop simple routines to allow researchers to take advantage of the information collected by smaller citizen science projects, for instance, programming specific packages to connect them to popular scientific environments (like R). Here, we present rAvis, an R-package to connect R-users with Proyecto AVIS (<http://proyectoavis.com>), a Spanish citizen science project with more than 82,000 bird observation records. We develop several functions to explore the database, to plot the geographic distribution of the species occurrences, and to generate personal queries to the database about species occurrences (number of individuals, distribution, etc.) and birdwatcher observations (number of species recorded by each collaborator, UTM's visited, etc.). This new R-package will allow scientists to access this database and to exploit the information generated by Spanish birdwatchers over the last 40 years.

Citation: Varela S, González-Hernández J, Casabella E, Barrientos R (2014) rAvis: An R-Package for Downloading Information Stored in Proyecto AVIS, a Citizen Science Bird Project. PLoS ONE 9(3): e91650. doi:10.1371/journal.pone.0091650

Editor: Stephanie Ann White, UCLA, United States of America

Received: November 24, 2013; **Accepted:** February 12, 2014; **Published:** March 13, 2014

Copyright: © 2014 Varela et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Sara Varela and Rafael Barrientos have postdoctoral grants. SV is funded by the European Science Foundation and Czech Republic, from the Education for Competitiveness Operational Programme (ECOP) project "Support of establishment, development and mobility of quality research teams at the Charles University" (CZ.1.07/2.3.00/30.0022)(<http://www.cuni.cz/UKEN-88.html>). RB received financial assistance from Juan de la Cierva program from the Spanish Ministry of Competitiveness (<http://www.idi.mineco.gob.es/portal/site/MICINN/menuitem.dbc68b34d11c5d52ffe801432ea0/?vgnnextoid=fdf669dad3f88310VgnVCM1000001d04140aRCRD>). Javier González-Hernández and Eduardo Casabella-Cabana are freelance web-developers that worked on the project during their free time. They were involved in the design, programming and optimization of the R-package ravis and on the web interface of Proyectoavis. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: svarela@paleobiogeography.org

Introduction

During the past several decades, developers have focused their attention on constructing web repositories to store and share biological information. On the one hand, there are online repositories with information generated by scientists, like specimens collected for museums and herbariums, fossil records or genetic data (e.g. GBIF: <http://gbif.org>, NOW: <http://helsinki.fi/science/now/>, GeneBank: <http://ncbi.nlm.nih.gov/genbank/>). On the other hand, there are web sites that store biological information collected by non-scientists, or so-called 'citizen science'.

Citizen science has proven to be an appropriate method to provide researchers with valuable information [1–3], and is increasingly used as an adequate way to sample species occurrences and distributions [4], to collect data to investigate urban ecology [3,5,6], or to collect data on bird biology, ecology and diversity [7–9]. In our case, data stored in Proyecto AVIS, our citizen science project to collect data from amateur Spanish ornithologists, show the same general patterns described by scientists based on their own samples and field experiments. Power law distributions of species/area [10] and species/abundance [11] have been detected (Figure 1), suggesting that

the data stored in Proyecto AVIS have similar properties to the data collected by scientists.

One of the main characteristics of the citizen science databases is that they are huge. For instance, birdwatchers' observations stored in the eBird database reached 100,000,000 observations and over 10,000 species (<http://ebird.org>). As a result, there are terabytes of information about species occurrences (latitude, longitude, altitude, time, habitat, diet, alleles, etc.) stored in online databases that follow different formats and standards of data storage [12], and the challenge now is developing easy strategies to use this information for research [13].

Currently, there are ongoing projects to generate tools to standardize the information stored in those databases (e.g. <http://ecodataretriever.org>) and to develop R-packages to connect online biological databases to the R-environment (<http://ropensci.org/>). As a consequence, large international databases are now being made available through R using packages like *rebird* [14], *rfishbase* [15], *rgbif* [16] or *rvertnet* [17] (connecting R with eBird, Fishbase, GBIF and VertNet databases, respectively). All of these new data exponentially increase our capabilities to answer questions about species conservation, global change, macroecology and biogeography.

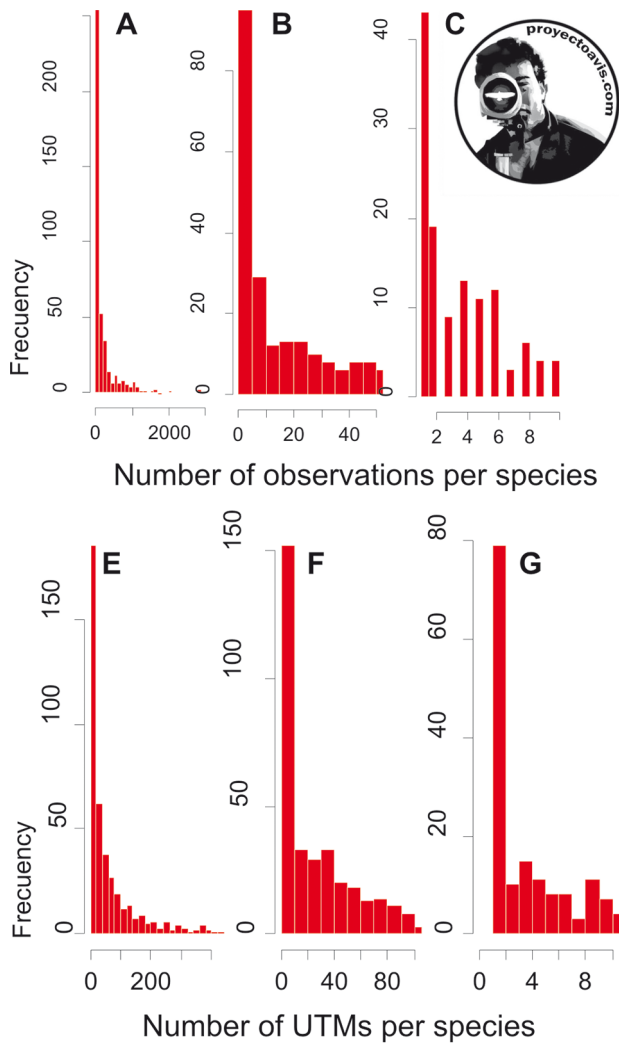


Figure 1. Data collected by amateur birdwatchers and stored in Proyecto AVIS show the same patterns as data collected by scientists, like scale invariant relationship of the frequency distribution of the number of observations per species (A-B-C) and scale invariant relationship of the frequency distribution of the number of UTMs per species (E-F-G).
doi:10.1371/journal.pone.0091650.g001

R is an open source and collaborative framework (<http://r-project.org/>), and is one of the most used environments for analyzing biological data and for developing scientific software [18]. Many young scientists are becoming advanced R-users (but see [19]). Thus, R is becoming a standard environment for developing easy-to-use (and re-use) functions and for sharing them with the academic community. For all of these reasons, we decided to build an R package to directly download the information stored in Proyecto AVIS from the R environment, in order to promote the use of the data stored in this database within the growing scientific R-community.

Proyecto AVIS

Each citizen science project stores singular and, consequently, important information [6,9,20–22]. Proyecto AVIS (<http://proyectoavis.com>) is a citizen science project born in August 2005 with the idea of collecting the data stored in the field notebooks of amateur Spanish ornithologists and sharing them

with both other amateur ornithologists and the scientific community. More than one hundred collaborators, including several NGOs, have been actively participating in the project uploading their bird observations. Overall, the database contains records over 40 years (1973–2013), stores 82,503 records, totalling 4,739,171 individuals from 413 species, which represents 90% of the total number of species recorded in Spain. In addition, it contains information from 1,717 different UTMs (squares of 10×10 km), representing 30% of the Spanish territory (query to the database: November 2013).

The Proyecto AVIS database and web page were built using open source software (MySQL, Perl, Apache) and free GIS layers. Proyecto AVIS requires five mandatory fields for each bird observation: ‘species’, ‘number of individuals’, ‘observation period’, ‘date’ and ‘UTM 10×10 km square’, plus several optional fields that include variables like ‘hour’, ‘sex’, ‘age’ or ‘habitat’. To standardize the taxonomy, the bird species list follows the Bird List of Spain from SEO/BirdLife [23]. Bird occurrences in the Proyecto AVIS database are georeferenced using the projected UTM 10×10 km square system and the MGRS labelling convention (Military Grid Reference System). The UTM/MGRS is the standard system for mapping species occurrences in Spain and is the system used by the Spanish bird atlases [24,25]. To help users identify the UTMs in which they recorded the species, the web application includes an easy-to-use tool to geo-referenced the observations based on a Google Maps™ routine.

The Proyecto AVIS web page (<http://proyectoavis.com>) includes several user-friendly tools for exploring the database, like summaries of the bird observations or graphics of the species records throughout the year, and it allows registered users to download detailed information about the species observations to Excel files. However, although the database is already available on the Internet, its use for research has not been properly exploited. Proyecto AVIS lacks a specific package to connect the web repository with the R-environment, and we believe that this fact has prevented scientists from using Proyecto AVIS information.

Description of the package

rAvis exclusively contains R code, which maximizes its portability across platforms, and it works in Unix-like and Windows operating systems. The rAvis functions have been optimized following the standards criteria for software quality [26,27] and they are accessible through GitHub (<https://github.com/javigzz/rAvis>). Bugs can be reported using GitHub; <https://github.com/javigzz/rAvis/issues>. rAvis is freely available on the Comprehensive R Archive Network; CRAN (<http://cran.r-project.org/>) and complete information about rAvis, its functions and their parameters is available in the package help.

rAvis uses functions from other R-packages to get and plot the data stored in Proyecto AVIS. Namely, R-libraries stringr [28], XML [29], tools [30], RCurl [31], scrapeR [32] and gdata [33] are used to download the bird observations; mapproj [34], raster [35] and rgdal [36] to plot the GIS files; and, finally, scales [37] is used to plot bird occurrences with a transparency.

Exploring Proyecto AVIS. We developed several functions to explore the database in an easy and visual way and other functions to download the selected information (see Table 1 and run the example). First, *avis.SpeciesSummary* allows users to download a table with a summary of the records stored in Proyecto AVIS aggregated by species: number of observations of each species, number of individuals recorded, number of different UTMs (10×10 km) with observations, number of birdwatchers that recorded the species. Second, *avis.ContributorsSummary* returns a table with a general summary of the records stored in the database

aggregated by birdwatcher: number of observations per birdwatcher, number of species observed, number of provinces with data, number of UTM's visited, number of periods of observations. Finally, *avisHasSpecies* checks if a species name exists in Proyecto AVIS and then, *avisMapSpecies* allows users to explore the distribution of the observations of the species by setting the name of the species and selecting the type of map; administrative boundaries ('admin') or physical map ('phys') (Figure 2).

For constructing the plots we used free GIS layers. We downloaded the Spanish administrative map from <http://diva-gis.org/>, the Spanish UTM map from the Spanish government online map repository http://bscw.rediris.es/pub/bscw.cgi/524254?client_size=1366x580, and the Spanish physical map from <http://openstreetmap.org/> using the R- library OpenStreetMap [38].

Advanced queries to Proyecto AVIS. We constructed two main functions to set flexible queries about the species occurrences and the birdwatcher observations: *avisQuerySpecies* and *avisQueryContributor*, respectively. These functions download the information stored in Proyecto AVIS, and are intended to be tuned by the users in relation to their specific objectives. Also, we programmed *avisQuery* as a flexible function to pass any argument allowed in Proyecto AVIS database. We decided not to predefine queries or to pre-process the data because this would narrow the possibilities for research [12]. Instead, we allow the users to set their own queries to Proyecto AVIS. Arguments include taxonomic levels, like species, family, order; individual characteristics, like age, sex, breeding status; temporal filters, like year and month; or environmental filters, like habitat. Moreover, we added a UTM-lalong conversion to all queries. Thus, the position of the observations is given in two different formats: projected UTM's 10x10 km and geographic coordinates WGS84 (common latitude-longitude coordinates, which are not available in the current web application from Proyecto AVIS). We did not program more specific graphics or statistical analyses because we understand that the purpose of this package is to obtain the biological information stored in Proyecto AVIS and not to re-program statistical

algorithms that are already available in other R-packages. We assume that R-users would employ different R-packages for calculating their own statistics and constructing their own plots (see the example).

Example. rAvis could be upgraded in future releases. To download the exact version of rAvis that we used in this example run the function *install_github* from *devtools* package as follows: `install_github("javigzz/rAvis", ref = "v0.1")`

Install rAvis from the CRAN and load the package `> install.packages("rAvis") > library(rAvis) > avisSetup(verbose = FALSE)` Check if the target species has records in Proyecto AVIS `> avisHasSpecies("Pica pica")` Plot the occurrences of the species to explore the data `> avisMapSpecies("Pica pica", matype = "phys")` Download the occurrences of the species `> Pica_pica <- avisQuerySpecies("Pica pica")` Filter the data using *avisQuery*. For instance, select only records from forests habitats setting `habitat = "bosque"` (the database is in Spanish) `> Pica_pica_forest <- avisQuery(species = "Pica pica", > habitat = "bosque")`

Plot the results using *avisMap* `> avisMap(Pica_pica_forest, label = "Pica pica; Forest")` If interested in several species, explore the database using *avisMapSpecies* `> avisMapSpecies(list("Tyto alba", "Athene noctua", > "Bubo bubo", "Strix aluco"), matype = "phys")` Save the maps individually using the *tiff* function `> directory <- "C:/your_directory" > species <- list("Tyto alba", "Athene noctua", > "Bubo bubo", "Strix aluco") > for(x in species){ > tiff(file.path(directory, paste("/", x, ".tiff", sep = ""))) > avisMapSpecies(x) > dev.off() > }`

Conclusions

We have programmed rAvis, an R-package designed to help researchers explore and download the information stored in Proyecto AVIS. Thus, biogeographers, macroecologists and ornithologists working in spatial ecology or temporal series, in addition to researchers working on citizen science can easily take advantage of the unique data stored in this database for their own research.

Table 1. Descriptions of the functions of the rAvis R-package.

Functions in rAvis	Description
<i>avisHasSpecies</i>	checks if a species name exists in Proyecto AVIS.
<i>avisSpeciesId</i>	Returns the id of the selected species
<i>avisAllSpecies</i>	Returns a list with the species names and the ids of the species in Proyecto AVIS
<i>avisMap</i>	Renders a map for the observations downloaded using <i>avisQuerySpecies</i>
<i>avisMapSpecies</i>	Renders a map for each of the selected species (could be a list of species)
<i>avisSpeciesSummary</i>	Downloads a table with a summary of the records stored in Proyecto AVIS aggregated by species; number of observations of each species, number of individuals recorded, number of different UTM's (10x10 km) with observations and number of birdwatchers that recorded the species
<i>avisQuery</i>	Downloads information using several filters, like Order, Family, Species, Age, Sex, Habitat, etc.
<i>avisQuerySpecies</i>	Wrapper for <i>avisQuery</i> that allows to perform a query for more than one species at once.
<i>avisContributorsSummary</i>	Returns a table with the observations aggregated by birdwatcher.
<i>avisContributorAggregatedObservations</i>	Downloads the information about the observations of a birdwatcher
<i>avisQueryContributor</i>	Wrapper for <i>avisQuery</i> that allows to perform a query for more than one contributor at once.
<i>avisSetup</i>	Allows the user to turn off the information messages provided by the functions using "verbose = FALSE"

doi:10.1371/journal.pone.0091650.t001

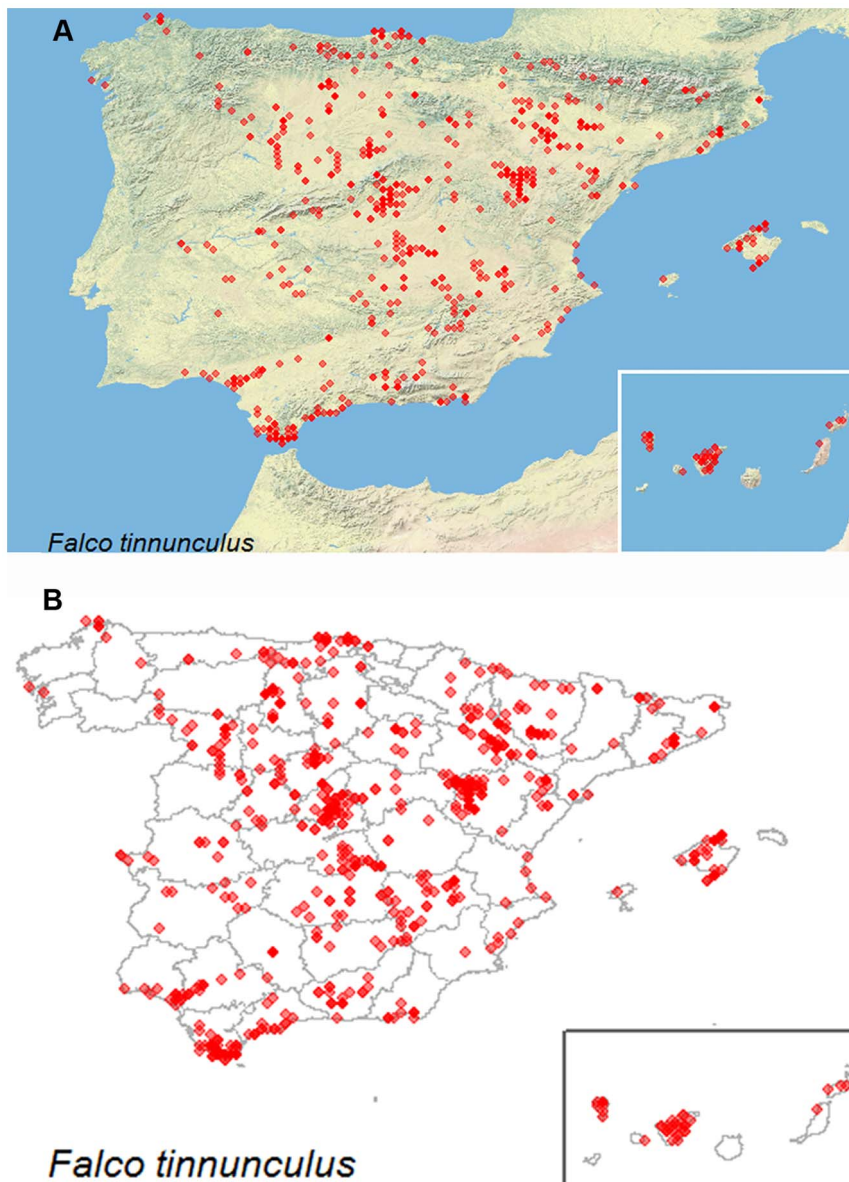


Figure 2. Outputs of the function *avisMapSpecies* setting the parameter *map* as 'phys' (A), or 'admin' (B) with the *Falco tinnunculus* records as an example.

doi:10.1371/journal.pone.0091650.g002

Acknowledgments

Thanks to Paco Montoya and the ornithologist association Cigüeña Negra from Tarifa (<http://cocn.tarifa.info.com>) for their important contribution to Proyecto AVIS. Special thanks to Jose Antonio Palomar and Juan Antonio Arce, who decisively contributed to Proyecto AVIS web development at several stages.

References

- Devictor V, Whittaker RJ, Beltrame C (2010) Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions* 16: 354–362.
- Oberhauser K, LeBuhn G (2012) Insects and plants: engaging undergraduates in authentic research through citizen science. *Frontiers in Ecology and the Environment* 10: 318–320.
- Nagy C, Bardwell K, Rockwell RF, Christie R, Weckel M (2012) Validation of a Citizen Science-Based Model of Site Occupancy for Eastern Screech Owls with Systematic Data in Suburban New York and Connecticut. *Northeastern Naturalist* 19: 143–158.
- Tulloch AIT, Possingham HP, Joseph LN, Szabo J, Martin TG (2013) Realising the full potential of citizen science monitoring programs. *Biological Conservation* 165: 128–138.
- Weckel ME, Mack D, Nagy C, Christie R, Wincorn A (2010) Using Citizen Science to Map Human-Coyote Interaction in Suburban New York, USA. *Journal of Wildlife Management* 74: 1163–1171.
- Mulder RA, Guay P-J, Wilson M, Coulson G (2010) Citizen science: recruiting residents for studies of tagged urban wildlife. *Wildlife Research* 37: 440–446.
- Cooper CB, Loyd KAT, Murante T, Savoca M, Dickinson J (2012) Natural History Traits Associated with Detecting Mortality Within Residential Bird

Author Contributions

Conceived and designed the experiments: SV JGH. Performed the experiments: JGH SV. Analyzed the data: SV. Contributed reagents/materials/analysis tools: EC RB. Wrote the paper: SV JGH EC RB.

- Communities: Can Citizen Science Provide Insights? *Environmental Management* 50: 11–20.
8. Wiersma YF (2010) Birding 2.0: Citizen Science and Effective Monitoring in the Web 2.0 World. *Avian Conservation and Ecology* 5: 13.
 9. Lepeczyk CA (2005) Integrating published data and citizen science to describe bird diversity across a landscape. *Journal of Applied Ecology* 42: 672–677.
 10. Sizing AL, Kunin WE, Sizingova E, Reif J, Storch D (2011) Between Geometry and Biology: The Problem of Universality of the Species-Area Relationship. *American Naturalist* 178: 602–611.
 11. Tjørve E, Kunin WE, Polce C, Calf Tjørve KM (2008) Species–area relationship: separating the effects of species abundance and spatial distribution. *Journal of Ecology* 96: 1141–1151.
 12. White EP, Baldrige E, Brym ZT, Locey KJ, McGlenn DJ, et al. (2013) Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution* 6: 1–10.
 13. Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and Opportunities of Open Data in Ecology. *Science* 331: 703–705.
 14. Maia R, Chamberlain S (2012) rebird: Interface to eBird. R package version 0.1. Available: <http://CRAN.R-project.org/package=rebird>.
 15. Boettiger C, Lang DT, Wainwright PC (2013) rfishbase: R Interface to FishBASE. R package version 0.2-1. Available: <http://CRAN.R-project.org/package=rfishbase>.
 16. Chamberlain S, Boettiger C, Ram K, Barve V (2013) rgbif: Interface to the Global Biodiversity Information Facility API methods. R package version 0.3.0. Available: <http://CRAN.R-project.org/package=rgbif>.
 17. Chamberlain S, Barve V (2012) rvertnet: Search VertNet database from R. Available: <http://CRAN.R-project.org/package=rvertnet>.
 18. Duck G, Nenadic G, Brass A, Robertson DL, Stevens R (2013) bioNerDS: exploring bioinformatics' database and software use through literature mining. *BMC Bioinformatics* 14.
 19. Joppa LN, McNerney G, Harper R, Salido L, Takeda K, et al. (2013) Troubling Trends in Scientific Software Use. *Science* 340: 814–815.
 20. Moyer-Horner L, Smith MM, Belt J (2012) Citizen science and observer variability during American pika surveys. *Journal of Wildlife Management* 76: 1472–1479.
 21. Worthington JP, Silvertown J, Cook L, Cameron R, Dodd M, et al. (2012) Evolution MegaLab: a case study in citizen science methods. *Methods in Ecology and Evolution* 3: 303–309.
 22. Silvertown J (2009) A new dawn for citizen science. *Trends in Ecology & Evolution* 24: 467–471.
 23. Gutiérrez R, de Juana E, Lorenzo JA (2012) Lista de las aves de España Edición de 2012. Sociedad Española de Ornitología (SEO/BirdLife).
 24. Martí R, Moral JCd (2002) Atlas de las aves reproductoras de España. Madrid: Dirección General de Conservación de la Naturaleza - Sociedad Española de Ornitología.
 25. SEO/BirdLife (2012) Atlas de las aves en invierno en España 2007–2010. Madrid: Ministerio de Agricultura, Alimentación y Medio Ambiente-SEO/BirdLife.
 26. Chambers JM (2008) Software for Data Analysis. Programming in R. Springer.
 27. Voulgaropoulou S, Spanos G, Angelis L, Ieee (2012) Analyzing Measurements of the R Statistical Open Source Software. Proceedings of the 2012 Ieee 35th Software Engineering Workshop. pp. 1–10.
 28. Wickham H (2012) stringr: Make it easier to work with strings. R package version 0.6.2. <http://CRAN.R-project.org/package=stringr>.
 29. Lang DT (2013) XML: Tools for parsing and generating XML within R and S-Plus. R package version 3.98-1.1. <http://CRAN.R-project.org/package=XML>.
 30. R-Core-Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://R-project.org/>.
 31. Lang DT (2013) RCurl: General network (HTTP/FTP/...) client interface for R. R package version 1.95-4.1. <http://CRAN.R-project.org/package=RCurl>.
 32. Acton RM (2010) scrapeR: Tools for Scraping Data from HTML and XML Documents. R package version 0.1.6. <http://CRAN.R-project.org/package=scrapeR>.
 33. Warnes GR, Bolker B, Gorjanc G, Grothendieck G, Korošec A, et al. (2013) gdata: Various R programming tools for data manipulation. R package version 2.13.2. <http://CRAN.R-project.org/package=gdata>.
 34. Bivand R, Lewin-Koh N (2013) maptools: Tools for reading and handling spatial objects. R package version 0.8–27. <http://CRAN.R-project.org/package=maptools>.
 35. Hijmans RJ (2013) raster: Geographic data analysis and modeling. R package version 2.1-62/r2833. <http://R-Forge.R-project.org/projects/raster/>.
 36. Bivand R, Keitt T, Rowlingson B (2013) rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8–11. <http://CRAN.R-project.org/package=rgdal>.
 37. Wickham H (2012) scales: Scale functions for graphics. R package version 0.2.3. <http://CRAN.R-project.org/package=scales>.
 38. Fellows I, Stotz J P (2013) OpenStreetMap: Access to open street map raster images. R package version 0.3.1. <http://CRAN.R-project.org/package=OpenStreetMap>.