Compartiendo datos en Ecología: como añadir (¿más?) valor a tus datos

Antonio Jesús Pérez-Luque1, Andrea Ros Candeira1, Francisco Rodríguez-Sánchez2

1. Laboratorio de Ecología, Instituto Interuniversitario de Investigación del Sistema Tierra (IISTA-CEAMA), Universidad de Granada, Avda. del Mediterráneo s/n, Granada 18006, España.
2. COMPLETAR

Autor para correspondencia: A. J. Pérez-Luque [[ajperez@ugr.es](mailto:ajperez@ugr.es)]

# Palabras clave

publicación de datos; repositorios de datos; data papers; metadatos

# Keywords

data publishing; data repository; data papers; metadata

IDEAS:

* Importancia de depositar los datos en repositorios institucionales o reconocidos por la comunidad científica
* Añadir valor añadido a los conjuntos de datos con la documentación de los mismos (metadatado) y la creación de DataPapers
* Enumerar algunas de las principales herramientas / plataformas / etc para documentación y depósito de datos y metadatos
* Inclusión de esta filosofía dentro del ciclo de Reproducibilidad en Ciencia
* Mostrar un caso de ejemplo con SINFONEVADA

Existen muchos repositorios de datos (Yan y Gerstein, 2011).

Un ejemplo.

## Como incluir ejemplos

Ajustamos un modelo lineal:

Utilizamos R (**???**) y Rmarkdown (**???**; **???**; **???**; **???**) para todos nuestros análisis. Para ajustar los modelos mixtos utilizamos lme4 (**???**).

# Resultados

Esta sección está dividida en subsecciones.

## Subsección 1

Los árboles de la parcela A fueron más altos que en la parcela B (altura media: 25 vs 13 m). Y muchos más resultados que se actualizan dinámicamente.

## Subsección 2

Texto.

## Subsección 3

Texto.

# Discusión

## Ejemplo SINFONEVADA

Las plantaciones de pinares presentan menor riqueza de especies y menor diversidad de especies que los bosques naturales de encinar y robledal. Esto se debe, en ambas variables, al menor múmero de especies herbáceas presentes en las plantaciones de pinar (**???**).

Para establecer los valores iniciales de riqueza en cada uno de los tipos de cobertura vegetal vamos a analizar los datos de campo procedentes de los inventarios forestales SINFONEVADA (**???**). Para ello consultaremos este [conjunto de datos](http://www.gbif.org/dataset/db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472) en GBIF donde está documentado.

El objetivo es ***obtener una tabla de riqueza por ecosistema***. Los pasos que realizaremos son:

* Estudiar el conjunto de datos. Consultar el identificador único (uuid) del dataset: db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472
* Descargar el dataset
* Obtener los valores de riqueza por ecosistema. Para ello utilizaremos el mapa de ecosistemas generado por el [Observatorio de Cambio Global de Sierra Nevada](www.obsnev.es) (**???**)

### Obtener el conjunto de datos

library("knitcitations")  
library("rgbif")  
library("rgdal")

## Loading required package: sp

## rgdal: version: 1.2-16, (SVN revision 701)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 2.1.2, released 2016/10/24  
## Path to GDAL shared files: /Users/ajpelu/Library/R/3.3/library/rgdal/gdal  
## GDAL binary built with GEOS: FALSE   
## Loaded PROJ.4 runtime: Rel. 4.9.1, 04 March 2015, [PJ\_VERSION: 491]  
## Path to PROJ.4 shared files: /Users/ajpelu/Library/R/3.3/library/rgdal/proj  
## Linking to sp version: 1.2-5

library("sp")  
library("raster")  
library("dplyr")

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:raster':  
##   
## intersect, select, union

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library("sf")

## Linking to GEOS 3.4.2, GDAL 2.1.2, proj.4 4.9.1

library("knitr")

En primer lugar tenemos que conocer el identificador único del conjunto de datos (UUID, *Universally Unique IDentifier*). En este caso, los datos se encuentran en alojados en GBIF, y se corresponde con el Data-Set key “db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472”, que se puede obtener de la pagina web del recurso de datos <https://www.gbif.org/dataset/db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472>.

Seguidamente obtenemos el conjunto de datos y sus metadatos, que coinciden con los metadatos incluidos en el repositorio donde se alojan los datos.

### UUID del conjunto de datos   
sinfo\_uuid <- 'db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472'  
  
## Metadatos   
sinfo\_meta <- datasets(uuid = sinfo\_uuid)

Podemos obtener el número total de ocurrencias que contiene el conjunto de datos, para lo cual escribimos

occ\_count(datasetKey=sinfo\_uuid)

## [1] 7920

Y seguidamente vamos a conseguir una tabla que contenga todas las ocurrencias del conjunto de datos

sinfonevada <- occ\_data(datasetKey=sinfo\_meta$data$key, limit = 8000)

### Numero de plots y especies por plot

* Obtendremos el numero de plots (agrupando por localización)
* Para cada plot calcularemos la riqueza de especies

# Get only the fields of interest   
df <- sf$data %>% dplyr::select(decimalLatitude, decimalLongitude, scientificName)  
  
# How many species by plot  
richness\_loc <- df %>%  
 group\_by(decimalLatitude, decimalLongitude) %>%   
 count() %>%   
 as.data.frame() %>%   
 tibble::rownames\_to\_column(var='id\_plot') %>%  
 rename(rich = n)

### Riqueza por ecosistemas

* Analizaremos la riqueza media de los plots por ecosistema
* Obtener los ecosistemas de OBSNEV, que están en este [enlace](https://www.dropbox.com/s/urcbafyczvqv7o1/ecosistemas_sn.zip?dl=0). También se han incluido en la carpeta /data del paquete.

# Prepare Ecosystems data   
wd <- getwd()  
  
temporalwd <- setwd(tempdir())  
unzip('../inst/extdata/ecosistemas\_sn.zip', exdir = temporalwd)  
  
eco <- readOGR(dsn=temporalwd, layer = 'ecosistemas', encoding="UTF-8", verbose = TRUE)  
  
# Transform projection  
eco\_t <- spTransform(eco, CRS("+init=epsg:4326"))

Ahora creamos una capa vectorial de puntos para los plots del invetnario forestal, y le asignamos su correspondiente proyección.

# Create an spatial point dataframe for the plots   
richness\_sp <- SpatialPointsDataFrame(richness\_loc[,c("decimalLongitude", "decimalLatitude")],  
 richness\_loc)  
projection(richness\_sp) <- CRS("+init=epsg:4326")

Seguidamente realizamos una intersección entre capas. Estamos interesados en obtener la tipología de ecosistema para cada plot del inventario forestal.

# See this example  
# https://gis.stackexchange.com/questions/226035/join-spatial-point-data-with-multiple-polygon-data-using-r  
  
# Convert to sf-objects  
richness\_sp.sf <- st\_as\_sf(richness\_sp)  
eco\_t.sf <- st\_as\_sf(eco\_t)  
  
# Keep all "meuse.sf", sort by row.names(meuse.sf). Default overlay is "intersects".  
aux <- st\_join(richness\_sp.sf, eco\_t.sf[,c('COD\_ECOSIS', 'ECOSISTE\_1')])  
  
# Convert back to Spatial\*  
richness\_sp\_eco <- as(aux, "Spatial")

Vamos a reagrupar los ecosistemas

aux <- aux %>% mutate(newECO = recode\_factor(COD\_ECOSIS,   
 `8`="Pine plantations",  
 `2`="High-mountain meadows",  
 `3`="High-mountain shrubland",  
 `5`="Mid-mountain shrubland",  
 `1`="Pastures",  
 `6`="Aquatic systems",   
 `NA`="NA",   
 .default = 'Natural Forests'))

Finalmente computamos los valores de riqueza por ecosistemas y por agregados

richSinfo <- aux %>%  
 group\_by(ECOSISTE\_1, COD\_ECOSIS) %>%  
 summarise(mean = mean(rich),  
 sd = sd(rich),  
 se = sd/sqrt(length(rich)),  
 n = length(rich),  
 min = min(rich),  
 max = max(rich),  
 median = median(rich)) %>%  
 as.data.frame() %>%  
 dplyr::select(-geometry)

richSinfo\_agg <- aux %>%  
 group\_by(newECO) %>%  
 summarise(mean = mean(rich),  
 sd = sd(rich),  
 se = sd/sqrt(length(rich)),  
 n = length(rich),  
 min = min(rich),  
 max = max(rich),   
 median = median(rich)) %>%  
 as.data.frame() %>%  
 dplyr::select(-geometry)

En (**???**) se analizan el mismo dataset, y se obtienen, tras aplicar los modelos, los siguientes resultados:

richPot <- data.frame(cbind(  
 eco = c('Plantations', 'Quercus ilex forests', 'Natural deciduous forests'),  
 n = c(442, 45, 26),  
 potRich = c(13.09, 14.92, 17.55),  
 lowerInterval = c(12.82, 13.72, 15.62),   
 upperInterval = c(13.34, 16.11, 19.66)))

# Conclusiones

Texto.

# Agradecimientos

LIFEADAPTAMED?  
ECOPOTENTIAL? Convenio OBSNEV?

###### REFERENCIAS

Yan, K.-K., Gerstein, M. 2011. The Spread of Scientific Information: Insights from the Web Usage Statistics in PLoS Article-Level Metrics Vespignani, A. (ed.), *PLoS ONE* 6: e19917.

###### TABLA 1

**Tabla 1**. Cada tabla debe aportar su correspondiente encabezamiento explicativo. En los Artículos de investigación, de revisión y Comunicaciones breves se aportarán los encabezamientos tanto en castellano como en inglés, en letra Arial 10 y en página independiente. Es importante que sean simples y que no superen el ancho una página DIN A4 vertical. Los originales se deben aportar en formato tabla y no en formato imagen.

**Table 1**. Table heading in English.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | setosa |
| 5.0 | 3.6 | 1.4 | 0.2 | setosa |
| 5.4 | 3.9 | 1.7 | 0.4 | setosa |

###### PIES DE FIGURA

**Figura 1**. Pie de figura 1.

**Figura 2**. Pie de figura 2.

###### FIGURE LEGENDS

**Figure 1**. Figure caption.

**Figure 2**. Figure caption.

###### FIGURA 1

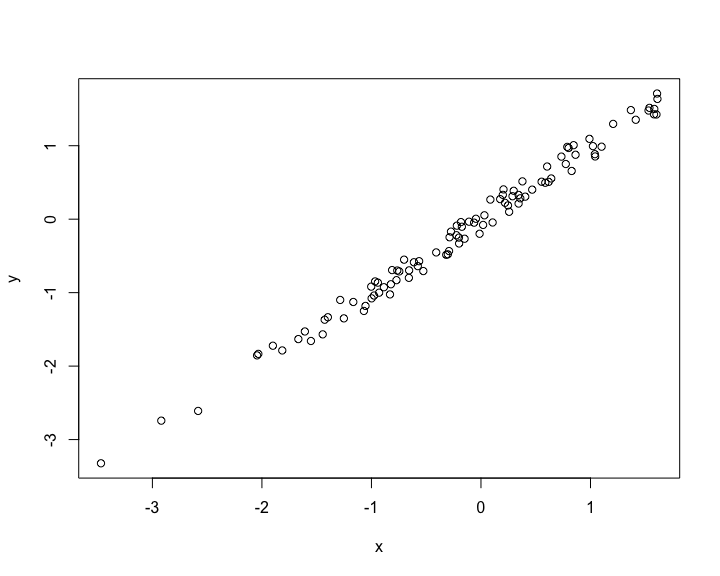


Figura 1. Esto es un ejemplo.

###### FIGURA 2

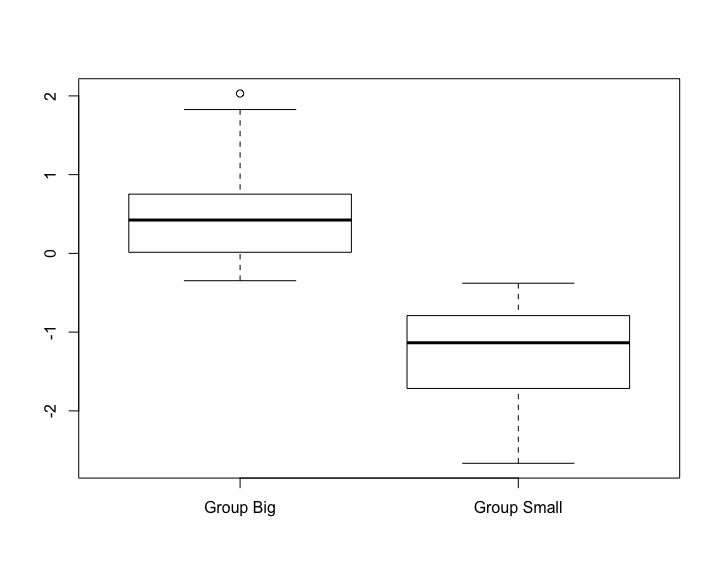


Figura 2. Segundo ejemplo