

**LATINR**

Conferencia Latinoamericana  
sobre el uso de R  
en Investigación + Desarrollo

# **Soy naturalista y quiero pasear en mi país, ¿dónde hay más oportunidades de llenar vacíos de información?**

Florencia Grattarola (@flograttarola)  
y Juan Manuel Barreneche  
(@jumanbar)

Latin R 2021

**¿Por qué hacer este proyecto?**

# Datos abiertos en un contexto de crisis de la biodiversidad

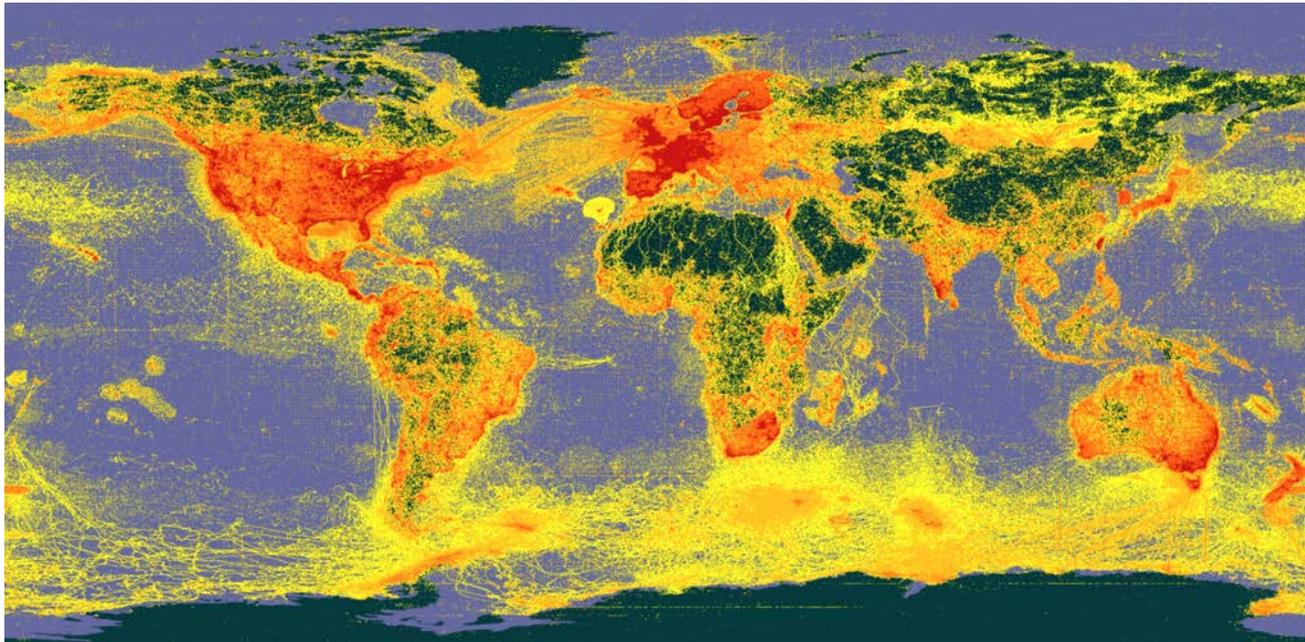
La biodiversidad a nivel global está disminuyendo a un ritmo sin precedentes.

Para poder evaluar estos cambios hay cada vez más datos disponibles abiertamente en todo el mundo.

- Digitalización de especímenes de museo
- Cámaras trampa
- Registradores de datos acústicos (e.j. Audiomoth)
- ADN ambiental
- Ciencia comunitaria



**Sin embargo, en diversas  
regiones de Latinoamérica y el  
Caribe los datos de biodiversidad  
disponibles de manera abierta  
son limitados**



Tal es el caso de Uruguay, en donde el 95% de su territorio permanece con muestreos insuficientes.

scientific reports

OPEN

## Multiple forms of hotspots of tetrapod biodiversity and the challenges of open-access data scarcity

Florencia Grattarola<sup>1,2,3</sup>, Juan A. Martínez-Lanfranco<sup>2</sup>, Germán Botto<sup>3</sup>, Daniel E. Naya<sup>4</sup>, Raúl Maneyro<sup>5</sup>, Patricia Mai<sup>6</sup>, Daniel Hernández<sup>7</sup>, Gabriel Laufer<sup>8</sup>, Lucía Ziegler<sup>8</sup>, Enrique M. González<sup>9</sup>, Inés da Rosa<sup>10</sup>, Noelia Gobel<sup>8</sup>, Andrés González<sup>2</sup>, Javier González<sup>2</sup>, Ana L. Rodales<sup>8</sup> & Daniel Pincheira-Donoso<sup>11,12</sup>

The uneven spatial distribution of biodiversity is a defining feature of nature. In fact, the implementation of conservation actions both locally and globally has progressively been guided by the identification of biodiversity 'hotspots' (areas with exceptional biodiversity). However, different regions of the world differ drastically in the availability of fine-scale data on the diversity and distribution of species, thus limiting the potential to assess their local environmental priorities. Within South America—a megadiverse continent—Uruguay represents a peculiar area where multiple tropical and non-tropical eco-regions converge, creating highly heterogeneous ecosystems, but where the systematic quantification of biodiversity remains largely anecdotal. To investigate the constraints posed by the limited access to biodiversity data, we employ the most comprehensive database for tetrapod vertebrates in Uruguay (spanning 664 species) assembled to date, to identify hotspots of species-richness, endemism and threatened species for the first time. Our results reveal negligible spatial congruence among biodiversity hotspots, and that tetrapod sampling has historically concentrated in only a few areas. Collectively, our study provides a detailed account of the areas where urgent biodiversity monitoring efforts are needed to develop more accurate knowledge on biodiversity patterns, offering government and environmental bodies a critical scientific resource for future planning.



# Datos abiertos en un contexto de crisis de la biodiversidad

Para poder tomar mejores decisiones basadas en evidencia es sumamente crítico revertir la falta de datos primarios sobre la distribución geográfica de las especies.

Para esto, la **ciencia comunitaria** se presenta como una herramienta colectiva transformadora.

# Ciencia comunitaria

Investigación **comunitaria** mediante la recolección, el análisis y la interpretación de datos, y la **integración de saberes**, para la generación de **nuevos conocimientos** y la creación de **oportunidades de aprendizaje**.



[www.inaturalist.org](http://www.inaturalist.org)

# iNaturalist

Esta plataforma web y aplicación para celulares, tiene como principal funcionalidad el registro de **organismos** en el **tiempo** y el **espacio** y reúne, en torno a estos datos, a la comunidad de naturalistas más grande del mundo.

Busca:

- Acercar a las personas a la naturaleza, su observación y conocimiento.
- Contribuir con la generación de datos de biodiversidad de calidad para su uso por parte de diferentes comunidades.

# Sesgos observacionales

Los datos que provienen de la ciencia comunitaria suelen originarse en áreas de fácil acceso y próximas a centros poblados, carreteras y áreas de interés (como, por ejemplo, áreas protegidas).

Para poder generar datos más diversos y mejorar el conocimiento de la biodiversidad, quienes observan (y quienes usan los datos) se verían muy beneficiados de tener a disposición una herramienta que les permita decidir **a dónde ir a registrar organismos y qué grupos observar en función de maximizar su aporte.**



# Objetivo

# Objetivo

Generar un mapa interactivo del Uruguay que categorice áreas según su **déficit de datos de biodiversidad** y nos permita resaltar aquellas en las que registros adicionales de biodiversidad podrían ser particularmente valiosos para **llenar los vacíos de conocimiento**.

**¿Cómo lo hicimos?**

# Datos de biodiversidad

Usamos como base los datos ingresados en la plataforma iNaturalist para Uruguay hasta el **21 de octubre de 2021**.

# Datos de biodiversidad

observed_on	license	url	taxon_id	scientific_name	common_name	class
2010-05-14	CC-BY-NC-ND	<a href="http://www.inaturalist.org/observations/25394">http://www.inaturalist.org/observations/25394</a>	10254	Paroaria coronata	Cardenal Copete Rojo	Aves
2010-05-17	CC-BY-NC-ND	<a href="http://www.inaturalist.org/observations/25695">http://www.inaturalist.org/observations/25695</a>	41736	Otariidae	Leones marinos y parientes	Mammalia
2010-05-13	CC-BY-NC-ND	<a href="http://www.inaturalist.org/observations/29269">http://www.inaturalist.org/observations/29269</a>	4411	Larus atlanticus	Gaviota cangrejera	Aves
2011-05-16	CC-BY-NC-ND	<a href="http://www.inaturalist.org/observations/29270">http://www.inaturalist.org/observations/29270</a>	4840	Haematopus palliatus	Ostrero americano	Aves
2010-05-14	CC-BY-NC-ND	<a href="http://www.inaturalist.org/observations/29274">http://www.inaturalist.org/observations/29274</a>	513889	Aramides cajaneus	Chiricote	Aves
2011-01-06	CC-BY-NC	<a href="http://www.inaturalist.org/observations/87220">http://www.inaturalist.org/observations/87220</a>	67108	Rhinella dorbignyi	Sapito Jardín	Amphibia
2011-01-06	CC-BY-NC	<a href="http://www.inaturalist.org/observations/87221">http://www.inaturalist.org/observations/87221</a>	24363	Scinax granulatus	Rana roncadora	Amphibia
2011-01-08	CC-BY-NC	<a href="http://www.inaturalist.org/observations/87222">http://www.inaturalist.org/observations/87222</a>	41667	Procyon cancrivorus	Mano pelada	Mammalia
2011-01-10	CC-BY-NC	<a href="http://www.inaturalist.org/observations/87224">http://www.inaturalist.org/observations/87224</a>	715178	Bunodosoma cangicum	NA	Anthozoa

# Datos de biodiversidad

```
iNatUY_filtro <- iNatUY %>%  
  dplyr::filter(captive_cultivated == FALSE,  
               coordinates_obscured == FALSE,  
               quality_grade=='research',  
               !is.na(taxon_species_name) & !is.na(icons_taxo
```

Eliminamos aquellos registros:

- de organismos **cautivos** o **cultivados**
- con **coordenadas oscurecidas**

# Datos de biodiversidad

```
iNatUY_filtro <- iNatUY %>%  
  dplyr::filter(captive_cultivated == FALSE,  
               coordinates_obscured == FALSE,  
               quality_grade=='research',  
               !is.na(taxon_species_name) & !is.na(icons_taxo
```

Y nos quedamos con los registros:

- identificados con **Grado de investigación**
- identificados a nivel de **especie**.

# Datos de biodiversidad

```
nrow(iNatUY) # datos totales
```

```
## [1] 28215
```

```
nrow(iNatUY_filtro) # datos analizados
```

```
## [1] 17398
```

**Porcentaje de datos retenidos sobre el total de datos:**

```
## [1] 61.66224
```

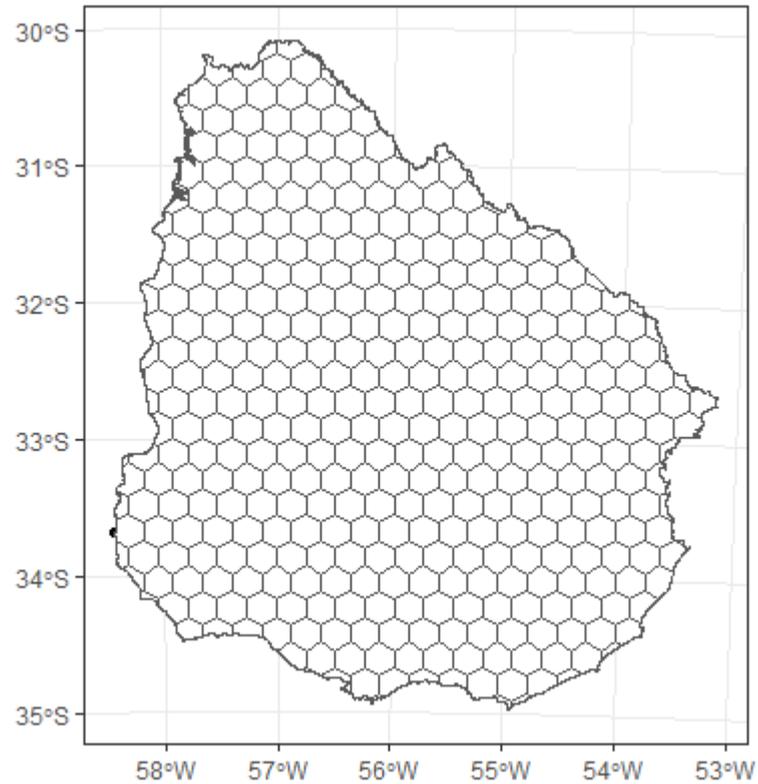
# Mapa base y celdas

Descargamos el mapa de Uruguay usando el paquete `geouy`.

```
Uruguay <- geouy::load_geouy(c = "Dptos")
```

Y luego dividimos el país en celdas hexagonales de  $500 \text{ km}^2$  usando los paquetes `sf` y `dplyr`.

```
grid_Uruguay <-  
  sf::st_make_grid(x = Uruguay, cellsize = 24028.11413  
  sf::st_intersection(., sf::st_union(Uruguay)) %>% sf::s  
  dplyr::mutate(grid_id = 1:nrow(.), area = sf::st_area(x
```



Celdas de 500  $km^2$

# Métricas utilizadas

Calculamos una serie de métricas por celda, usando el paquete sf:

- **Intensidad espacial**
- **Intensidad temporal**

Y luego las integramos en un:

- **Índice de Prioridad**

# Intensidad espacial

La cantidad de registros por unidad de área.

$$IE = n_r / A_c$$

En donde  $n_r$  y  $A_c$  son la cantidad de registros y área para una celda determinada, respectivamente.

# Intensidad temporal

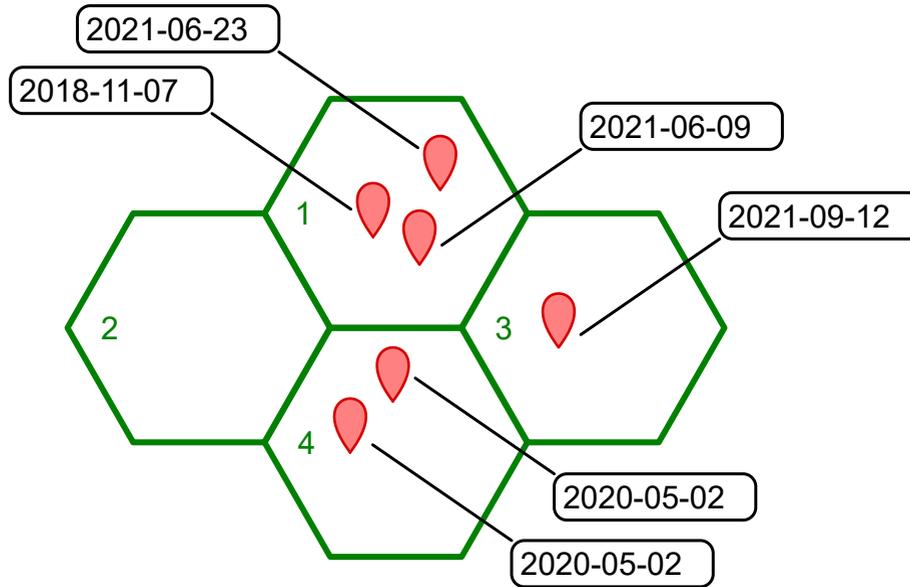
La cantidad de registros con **pares año-mes** diferentes.

$$IT = n_{\text{año.mes}}$$

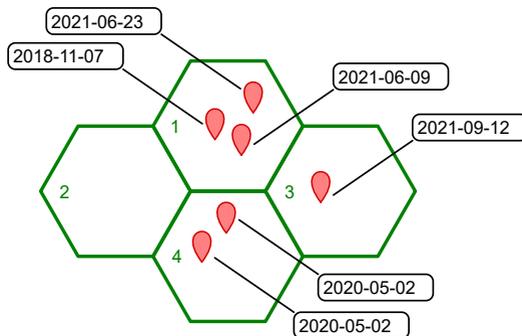
En donde  $n_{\text{año.mes}}$  es el número de registros con **pares año-mes** diferentes para una celda determinada.

Por ejemplo, las fechas 2021-08-09 y 2021-08-31, están en el mismo par año-mes, pero las fechas 2021-08-09 y 2021-09-01 no.

# Ejemplo

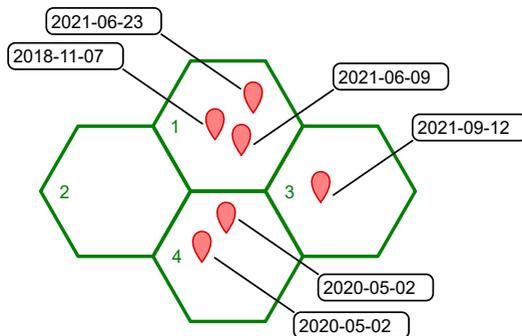


# Ejemplo



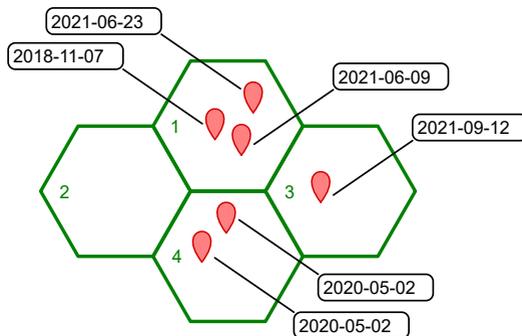
- En la celda 1, tenemos 3 registros, por lo que la IE será  $3/\text{área} = 0.3$  (área =  $10 \text{ km}^2$ ).
- En cambio la IT es 2, ya que hay 3 registros pero **sólo hay dos combinaciones año-mes: 2018-11 y 2021-06**

# Ejemplo



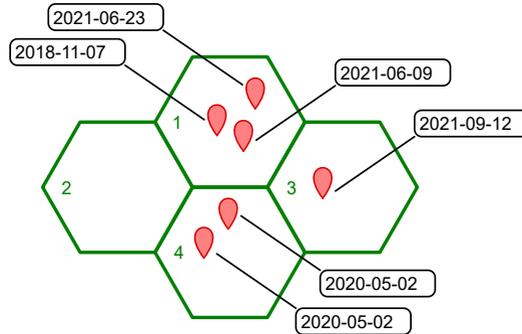
- En la celda 2 tenemos 0 registros, así que tanto IE como IT serán 0.

# Ejemplo



- En la celda 3 hay un registro, así que:  $IE = 0.1$  e  $IT = 1$

# Ejemplo



- En la celda 4 hay dos registros del mismo día (y por lo tanto, del mismo año-mes: 2020-05), por lo que  $IE = 0.2$  e  $IT = 1$ .

# Índice de Prioridad

La idea es combinar ambas medidas de intensidad para obtener un **índice de prioridad global**, que nos permita diferenciar celdas según sus relativas falta o presencia de datos.

# Índice de Prioridad

En nuestra app, el IP se calcula con la función `calc_ip`:

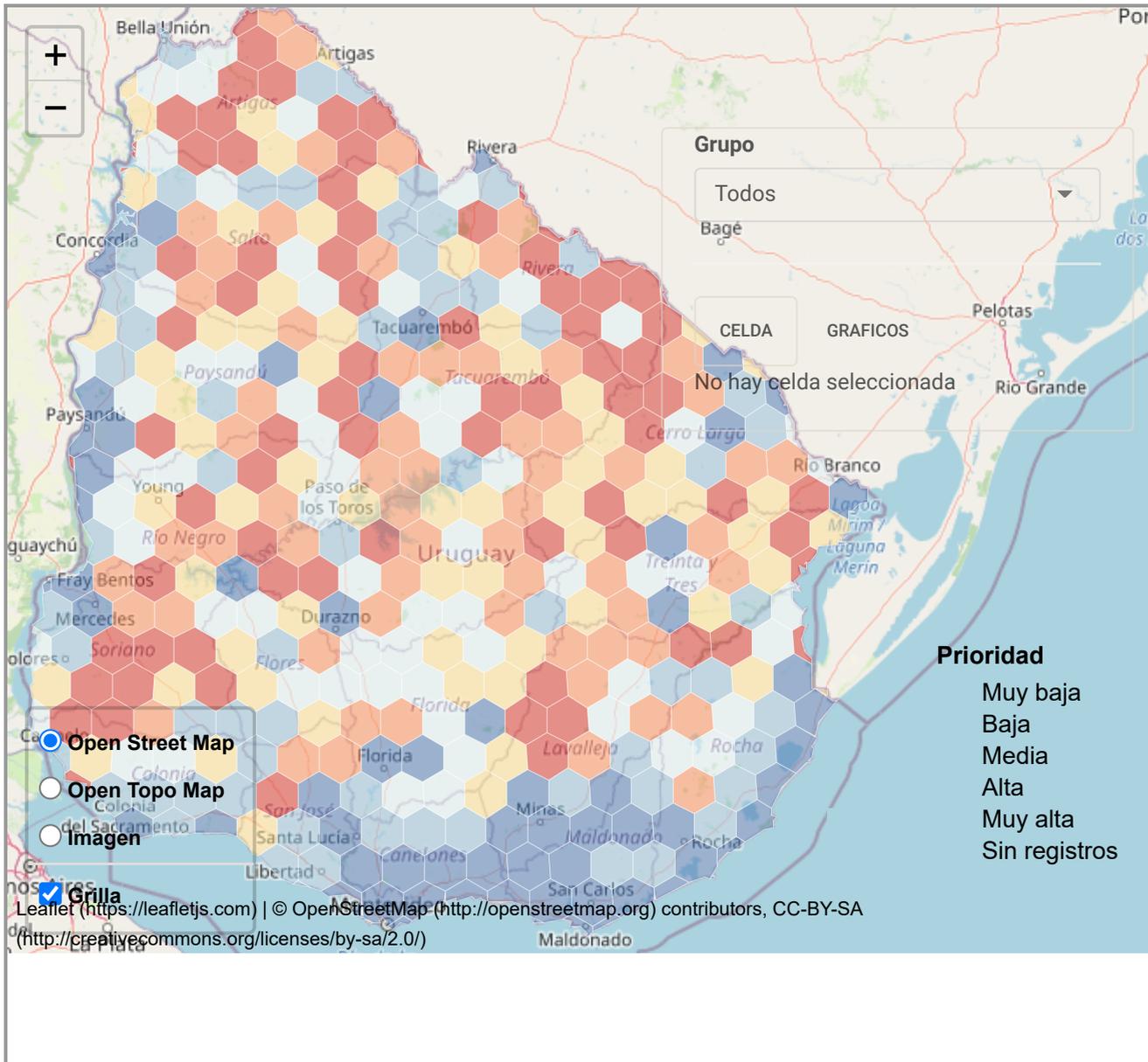
```
calc_ip <- function(it, ie, n.reg) {  
  # 1. rescalamientos:  
  it = scales::rescale(it, to = 0:1)  
  ie = scales::rescale(ie, to = 0:1)  
  # 2. suma:  
  suma <- it + ie  
  # 3. ranking  
  r <- rank(suma, ties.method = 'min', na.last = TRUE)  
  # 4. rescalamiento final (ranking a percentiles):  
  out <- scales::rescale(r, to = 1:0)  
  return(out)  
}
```

# Índice de Prioridad

Finalmente, las celdas fueron categorizados según el índice de prioridad, considerando a todas las especies juntas y cada grupo icónico.



# Resultados



Leaflet (<https://leafletjs.com>) | © OpenStreetMap (<http://openstreetmap.org>) contributors, CC-BY-SA  
 (<http://creativecommons.org/licenses/by-sa/2.0/>)

**A futuro**

# A futuro

Esta herramienta podría replicarse en otras partes del mundo, además, nuevas métricas podrían ser propuestas y nuevas formas de visualizar los datos acordes a las regiones (e.g. resaltando rutas, áreas administrativas o zonas de interés).

Esta es la primera vez que hacemos pública la herramienta por lo que esperamos poder recibir devoluciones para seguirla mejorando a futuro.

# ¡GRACIAS!

Florencia Grattarola ([@flograttarola](#))  
Juan Manuel Barreneche ([@jumanbar](#))



**JULANA**  
Jugando en la Naturaleza



Czech University  
of Life Sciences Prague



**NaturalistaUY**



# Referencias

# Referencias

- GBIF Secretariat (2021) Diapositivas de resumen de GBIF [Octubre 2021]. Copenhagen: GBIF Secretariat.  
<https://www.gbif.org/document/81771>.
- Grattarola F. et al. (2020) Multiple forms of hotspots of tetrapod biodiversity and the challenges of open-access data scarcity. Sci Rep 10, 22045.  
<https://doi.org/10.1038/s41598-020-79074-8>
- iNaturalist (2021) Datos descargados [21 Octubre 2021] [www.inaturalist.org](http://www.inaturalist.org)
- Wickham et al., (2019) Welcome to the tidyverse. Journal of Open Source

# Referencias

- Pebesma E. (2018) Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>
- Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2021). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.4.1. <https://CRAN.R-project.org/package=leaflet>
- Detomasi R. (2020) geouy: Geographic Information of Uruguay. R package version 0.2.2. <https://github.com/RichDeto/geouy>.

**¡Les invitamos a sumarse a  
iNaturalist para contribuir a  
mejorar el conocimiento  
sobre la biodiversidad en  
nuestra región!**