# Exportá los datos de tus plots con ggdatasaver

*Elio Campitelli, Elizabeth Hare*

**Palabras clave:** ggplot2, reproducibilidad, accesibilidad, rmarkdown

## Abstract

Es muy común leer un paper o publicación y querer reproducir parte del análisis, gráficos o realizar un análisis alternativo. Cuando los datos necesarios no están disponibles, muchas veces es necesario recurrir a la digitalización de gŕaficos. Mediante este proceso manual, lento, inexacto y propenso a errores se pueden obtener las coordenadas aproximadas asociadas a puntos o líneas de un gráfico.

Sin embargo, esto no sería necesario si los datos específicos mostrados en los gráficos estuvieran accesibles. El paquete **ggdatasaver** (github.com/eliocamp/ggdatasaver) guarda automáticamente las coordenadas de cada capa de cada gráfico creado con ggplot2. Los datos de cada gráfico se guardan en un archivo comprimido que contiene archivos csv de cada capa y uno archivo que define la disposición de los paneles. Los archivos luego se pueden compartir como material suplementario en una publicación o subir a un repositorio público como Zenodo o Figshare.

## Modo de uso

La mejor forma de usar ggdatasaver es cuando se utiliza dentro de un documento de rmarkdown. En este caso lo único que le usuarie tiene que hacer es determinar el directorio donde se van a guardar los archivos poniendo

```
ggdatasaver::save_plot_data_in("plot-data")
```

En el primer bloque de código del documento. Al knitear el documento, todos los plots que se muestren debajo de esa línea serán procesados y sus datos guardados en el directorio “plot-data” relativo al directorio de trabajo.

## Accesibilidad

Las revistas académicas casi nunca tienen una infraestructura que permita el texto alternativo para las figuras, pero sí para datos suplementarios. Para las personas ciegas, tener acceso a los datos crudos usados para generarlas es mejor que nada.

Teniendo acceso a estos datos, personas ciegas podrían generar tablas de datos y calcular estadísticas accesibles con con lector de pantallas para tener una mejor idea de las

relaciones subyacentes, o simplemente leer los números. En el caso de las curvas salidas de modelo, que normalmente no se describen adecuadamente en el texto, tener los datos permite poder realizar el ajuste y leer los parámetros del mismo.

## Reproducibilidad

Un aspecto importante de la reproducibilidad es tener acceso a los datos, pero esto no siempre es fácil o factible. Almacenar y transferir grandes cantidades de datos es caro, y muchos tipos de datos conllevan problemas de privacidad (como los datos médicos) o de licencias (como los datos obtenidos bajo licencias restrictivas). Otro obstáculo para compartir datos es organizarlos de forma útil.

Aunque no es perfecto, compartir los pequeños fragmentos de datos que son las coordenadas de las geometrías de las parcelas puede ser un buen compromiso. Estos datos son generalmente pequeños y ya están en un formato tabular, por lo que es técnicamente fácil de compartir en un repositorio o como material complementario. Y como son datos que ya se comparten implícitamente como imagen, los problemas de privacidad y licencias son mucho menores.

Pero aún en casos donde es posible compartir todos los datos crudos, compartir también los datos de los gráficos puede ser útil para investigadores que quieran reproducir o reanalizar pequeños trozos los resultados pero no quieran o no pueden descargar los datos originales y ejecutar el código.

## Conclusión

Compartir los datos asociados a cada figura tiene ventajas tanto en términos de accesibilidad como de reproducibilidad. ggdatasaver permite hacerlo sin necesidad de cambiar el código ya existente y con sólo una línea de código.

Elio Campitelli
Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos. Buenos Aires, Argentina.
eliocampitelli@gmail.com

Elizabeth Hare
Dog Genetics LLC
Astoria, NY, EEUU
lizhare@gmail.com