

De la zoología a las aplicaciones generales, *unheadR*: un paquete para estructurar datos no ordenados

Al momento de formatear, almacenar, y compartir datos con colaboradores o con el público en general, existen ciertas prácticas comunes que pueden generar obstáculos para el posterior uso de estos datos. Generalmente son cambios que hacemos con fines estéticos, o ajustes menores para minimizar el espacio que ocupa una tabla en una página impresa o digital. Aunque sean cambios menores, pueden modificar la estructura de los datos. Por ejemplo: al insertar subencabezados dentro de otras variables para delimitar grupos de observaciones (Figura 1), o partir en dos o más filas los valores de la variable que contiene los nombres de cada unidad observacional (Figura 2). Los datos compactos y fáciles de leer no siempre son fáciles de usar en programas de análisis, por lo que es recomendable siempre compartir tablas y conjuntos de datos en estructura ordenada (en la cual cada variable tiene su propia columna, cada observación tiene su propia fila, y cada valor tiene su propia celda).

Figura 1. a) Uso de subencabezados para agrupar observaciones, b) Valores de la variable de agrupación en su propia columna. Los datos fueron ordenados utilizando *unheadR* con una expresión regular para identificar las familias taxonomicas dentro la la variable de nombres científicos, identificadas por el sufijo en latín ‘dae’.

a)

| Nombre Científico | Edo. Conservación |
|-------------------------------|-------------------|
| Leporidae | |
| <i>Brachylagus idahoensis</i> | Estable |
| <i>Caprolagus hispidus</i> | Amenazado |
| Ochotonidae | |
| <i>Ochotona alpina</i> | Estable |
| <i>Ochotona iliensis</i> | Amenazado |
| <i>Ochotona princeps</i> | Estable |

b)

| Nombre Científico | Edo. Conservación | Familia |
|-------------------------------|-------------------|-------------|
| <i>Brachylagus idahoensis</i> | Estable | Leporidae |
| <i>Caprolagus hispidus</i> | Amenazado | Leporidae |
| <i>Ochotona alpina</i> | Estable | Ochotonidae |
| <i>Ochotona iliensis</i> | Amenazado | Ochotonidae |
| <i>Ochotona princeps</i> | Estable | Ochotonidae |

Cuando no hay opción y tenemos que trabajar con datos no ordenados que contienen subencabezados o valores rotos, es recomendable reorganizar estos datos de manera programática para ahorrar tiempo y además evitar introducir errores no reproducibles. Este trabajo describe el paquete *unheadR* (<https://github.com/luisDVA/unheadr/>), que incluye tres funciones simples que aprovechan las herramientas de ‘tidy evaluation’ para programar usando funciones de los paquetes *dplyr* y *tidyr* mediante evaluación de expresiones no estándar. El paquete fue diseñado para resolver estos obstáculos comunes en la estructura de los datos y generar conjuntos de datos ordenados. Las funciones fueron planeadas para usarse en el contexto de *tidyverse*, un conjunto de paquetes con una filosofía común y diseñados para utilizarse juntos.

Se describe el desarrollo y uso del paquete *unheadR* para la re-digitalización de cientos de tablas proporcionadas en artículos científicos sobre biología de mamíferos. Los subencabezados se utilizan ampliamente en ciencias biológicas para reportar grupos de especies que pertenecen a la misma familia o gremio ecológico, y representan un obstáculo significativo para la reutilización de datos taxonómicos. A partir de la implementación inicial del paquete, se definió una nomenclatura general de las diferentes estrategias para ordenar datos, y se delineó una estrategia general para aplicar estas funciones a todo tipo de datos.

Figura 2. a) Tabla de datos con valores rotos en la variable que contiene las unidades observacionales (especies de murciélago) y valores en blanco, b) Presentación ordenada de los mismos datos mediante las funciones de *unheadR*.

unidades observacionales

a)

| Especie focal | Dieta | Abundancia |
|--------------------|-------------|------------|
| Glossophaga | Nectarívoro | 71288 |
| soricina | NA | NA |
| Myotis flavus | Insectívoro | 5902 |
| Artibeus lituratus | Frugívoro | 4950 |

b)

| Especie focal | Dieta | Abundancia |
|----------------------|-------------|------------|
| Glossophaga soricina | Nectarívoro | 71288 |
| Myotis flavus | Insectívoro | 5902 |
| Artibeus lituratus | Frugívoro | 4950 |