Tópicos de interés:

Desarrollo de nuevos paquetes R

* Análisis de grandes datos con R
* Uso de R para análisis de datos abiertos

Idioma: español

<https://ropensci.org/blog/2024/02/15/champions-program-champions-2024/#andrea-gomez-vargas>

<https://ropensci.org/blog/2023/07/14/call-champions-program/>

ecosistema de la Ciencia Abierta y en la Comunidad R en particular,

Hemos desarrollado un programa de doce meses que incluye, capacitación basado en cohortes, el desarrollo de un proyecto y tutorías individuales.

crear un nuevo paquete,

{ARcenso}: Primeros pasos desarrollando un paquete en comunidad con rOpenSci

**Palabras clave:**  R, Shiny, paquete, censo, Estadísticas de población, Argentina, comunidad.

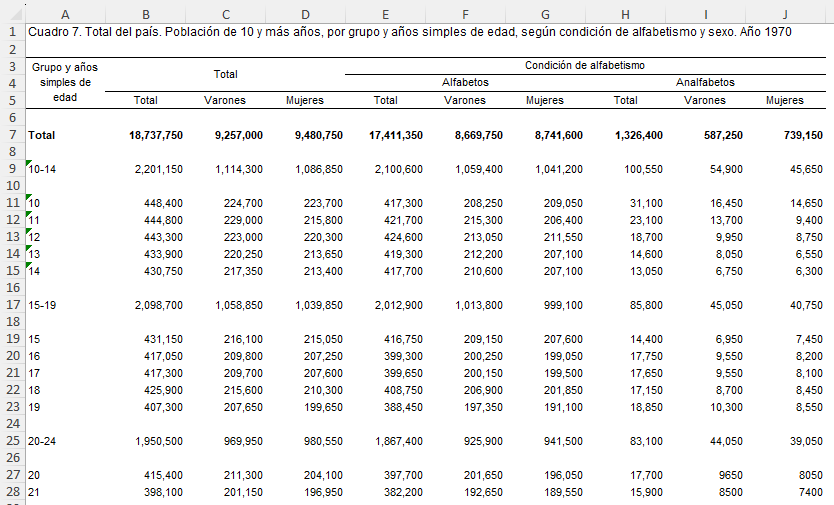
En el marco del programa de Campeones y Campeonas de rOpenSci[[1]](#footnote-0) y con el objetivo de identificar, reconocer y recompensar a los miembros apasionados de la comunidad que ayudan a la comunidad a crecer y mejorar nace el proyecto de crear el paquete {arcenso}.



{arcenso} es un paquete en desarrollo que permite acceder a los datos oficiales de los censos nacionales de población provenientes del Instituto Nacional de Estadísticas y Censos - INDEC desde 1970 a 2022.

Actualmente los resultados históricos de los censos nacionales de población están disponibles únicamente a través de archivos en formato excel no pensados para ser tratados o procesados como base de datos (Figura 1). Además a través de los censos la presentación de los datos no está homogeneizada dificultando la comparación a través del tiempo de la información disponible.

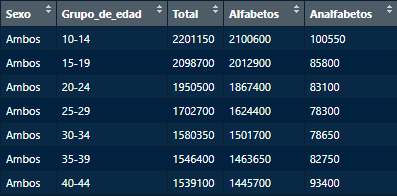
Figura 1. Información disponible en la web de INDEC.



En este marco contar con un paquete de información censal permitirá al sector público y privado, a los ciudadanos y a otros actores de la sociedad acceder a información actual e histórica sobre la población, los hogares y las viviendas de Argentina de una manera más accesible amplificando el ecosistema de la ciencia abierta.

Contar con un paquete con estas características requiere de un marco metodológico y conceptual implica la conversión de los tabulados publicados por INDEC a un formato “tidy” (Figura 2) preparado para ser utilizados de forma directa sin necesidad de realizar un pre-proceso de preparación de los datos por cada tabla que se quiera incorporar al análisis.

Figura 2. Mismo tabulado en formato “tidy”



El proyecto AR\_CENSO consiste en generar un paquete que permita disponer de datos censales homogeneizados y listos para su uso. Lo que implica no sólo la conversión de los tabulados publicados, sino también aportar herramientas en shiny que ayuden a la visualización de los datos y el armado de funciones para simplificar la comparabilidad y utilización conjunta de la información de múltiples años.

En esta presentación nos proponemos compartir reflexiones sobre el armado del proyecto,

la toma de decisio nes en la construcción de una visualización interactiva en Shiny y el desarrollo de un proyecto que permita a la ciudadanía y los usuarios de estadísticas públicas un mayor acercamiento a la producción pública de datos y el desarrollo de marcos conceptuales. También, nos interesa desarrollar los aprendizajes y enseñanzas durante este proceso a nivel colectivo y colaborativo, focalizando en la interacción no sólo con otros equipos al interior del organismo, sino también con otros organismos públicos y la comunidad de R. Por último, queremos reflexionar sobre los impactos del uso de software libre para la producción estatal de información estadística.

1. [¡Convocatoria abierta para la presentación de candidaturas al Programa de Campeonas y Campeones de rOpenSci!](https://ropensci.org/es/blog/2023/07/14/launch-champions-program-es/) [↑](#footnote-ref-0)