

# desuctools: paquete para para análisis de encuestas sociales

Anónimo

**Palabras clave:** encuestas - análisis reproducible - mejora de procesos - ponderación

La **Dirección de Estudios Sociales (DESUC)** es una unidad de estudios y servicios profesionales del Instituto de Sociología de la Pontificia Universidad Católica de Chile que trabaja en diversas áreas y materias sociales, de opinión pública, educación, comunicaciones, estudios organizacionales, entre otros temas. Para ello se ha especializado en el levantamiento y análisis de datos **cualitativos** y **cuantitativos**. En el ámbito cuantitativo realiza levantamiento de datos presenciales, telefónicos en un *call center* propio y por Internet.

## Contexto

Para el seguimiento, validación, análisis y comunicación de datos DESUC ha ocupado historicamente procesos y herramientas separadas, muchas veces manuales, las cuales podrían aumentar las posibilidades de cometer errores en el proceso. Otras implicancias de este tipo de prácticas fueron señaladas en la charla *Reproducible Analysis in the UK Government* presentada en LatinR 2019<sup>1</sup>. En particular, en el trabajo en DESUC destacamos tres problemas vinculados a ello:

- El creciente interés por análisis estadísticos sofisticados hacía más habitual el uso de SPSS para obtener frecuencias, y Stata para análisis de regresión y modelamientos. Esto podía generar bases de datos duplicadas con posibles imputaciones y codificaciones de variables divergentes. Por otro lado, para la generación de reportes el uso de Excel era habitual. Debido a ello, para supervisar y apoyar en estas diferentes tareas —además de orden— se necesitaba conocimiento de tres softwares distintos.
- Lo anterior repercutió en la velocidad con la que se puede procesar y analizar la información recolectada, sobre todo cuando existen procesos iterativos de revisión y ajuste de los datos.
- El principal efecto negativo de lo anterior es la lentitud con la que pueden plasmarse innovaciones y mejoras de procesos logrados en un estudio, en un proyecto futuro. Dado esto, es difícil tanto realizar procesos de análisis reproducibles, como también generar un *toolbox* en donde queden almacenadas ideas que puedan ser reutilizadas posteriormente en otros contextos.

## Solución

Para acumular, potenciar y hacer disponible el conocimiento acumulado a lo largo de los distintos proyectos que toca conducir, se decidió crear el paquete de R **desuctools**, el cual contiene funciones o datos que son utilizados habitualmente para la selección de muestras, análisis y visualización de datos de encuestas sociales de la manera idiosincrática en que DESUC lo hace.

El desarrollo de esta herramienta, además de potenciar y mejorar los aspectos recién señalados, ha tenido efectos de segundo orden dentro de la organización que vale la pena destacar. Primero, ha inspirado a detectar patrones habituales del trabajo realizado, y de esta manera, analizar cuales de ellos puede ser paquetizado en una función. Segundo, mostró la necesidad de utilizar sistemas de desarrollo colaborativo de código y control de versiones —GitHub en nuestro caso. Y tercero, permitió seleccionar y unificar los paquetes que habitualmente utilizamos internamente para analizar y visualizar datos y encuestas.

Finalmente, **desuctools** es un paquete que está en continuo desarrollo, tanto mejorando las funciones que tiene, como incluyendo nuevas según las nuevas ideas o necesidades de los proyectos y el equipo.

<sup>1</sup>Blog: <https://dataingovernment.blog.gov.uk/2017/03/27/reproducible-analytical-pipeline/>

## ¿Cuánto temor le producen las siguientes situaciones?

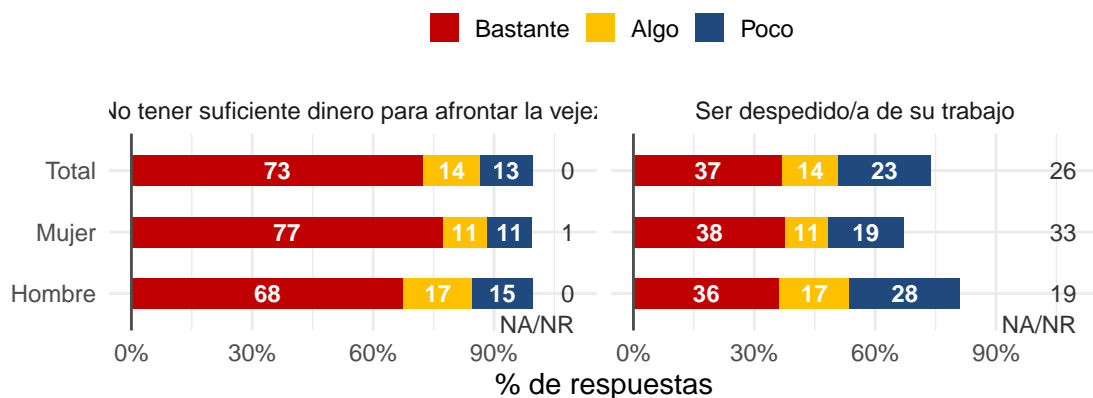


Figura 1: Gráfico de ejemplo

## Función destacada

Para esta presentación destacaremos la función `tabla_vars_segmentos`, la cual permite generar un `data.frame` *tidy* con la cantidad y proporción de respuestas para las categorías de un número arbitrario de preguntas para un número arbitrario de segmentos de la población. Las respuestas podrán ser ponderados según una variable auxiliar adicional, además de calcularse porcentajes válidos en el caso que se quiera establecer una o más categorías como *missing*. Adicionalmente, se generan columnas adicionales con etiquetas de variables y categorías para obtener una tabla en el formato adecuado para facilitar la generación de tablas y gráficos de resultados.

En el proceso de creación de la figura ??, se utilizan también funciones adicionales: `desuctools::rec_cat_5a3`, `desuctools::gg_bar_3_niveles_stack` y `desuctools::str_entre_parentesis`.

```
suppressMessages(library(tidyverse))
# install_github('desuc/desuctools')

file <- tempfile()
download.file(url = 'https://github.com/DESUC/30diasdegraficos/raw/master/inputs/12-lollipop-df_bicen_19_300',
              destfile = file)

data <- readRDS(file) %>%
  mutate(across(c(t01_1:t01_2),
    ~desuctools::rec_cat_5a3(., labels = c('Bastante' = 1, 'Algo' = 2, 'Poco' = 3, 'NA/NR' = 9)),
    total = TRUE, # Incluye dato total
    .vars = vars(t01_1, t01_2), # Listado de variables de interés
    .segmentos = vars(d07), # Listado de segmentos de interés
    .wt = pond_se) %>% # Ponderador
  mutate(pregunta_lab = desuctools::str_entre_parentesis(pregunta_lab))

desuctools::gg_bar_3_niveles_stack(
  .df = data,
  x = segmento_cat,
  facet_col = pregunta_lab,
  missing = 'NA/NR', y_na = 1.1, x_na = -2.5,
  title = '¿Cuánto temor le producen las siguientes situaciones?',
  font_family = '')
```