Datos etiquetados con labelled : : **CHEAT SHEET**

El paquete **labelled** ofrece funciones y métodos para manejar y manipular datos etiquetados, como los importados con **haven**.

Fundamentos

Los datos etiquetados son una estructura común en otros entornos estadísticos como Stata, SAS o SPSS.

Consisten en atributos adicionales de vectores numéricos y de caracteres (incluidas columnas de un data frame).

Hay 3 tipos de atributos:

1. **Etiquetas de variable** (descripción breve)
2. **Etiquetas de valor** (de valores específicos)
3. **Valores faltantes**:
   - Faltantes definidos por el usuario (estilo SPSS)
   - NA etiquetados (estilo Stata y SAS)

Etiquetas de variables

MANIPULAR UN VECTOR

var_label(x) *o* **var_label**(df$v1)
Obtiene la etiqueta de variable asociada a un vector x

var_label(x) <- "variable description"
Añade/modifica la etiqueta de variable de x

var_label(x) <- NULL
Elimina la etiqueta de variable asociada a x

MANIPULAR UN DATA.FRAME

var_label(df)
Lista todas las etiquetas de variable de las columnas de df

var_label(df) <- list(v1 = "variable 1", v2 = "variable 2")
Actualiza etiquetas de variable de columnas de df

df %>% **set_variable_labels**(v1 = "variable 1", v2 = "variable 2", v3 = NULL)
Actualiza las etiquetas de variable con la sintaxis de dplyr

df %>% **look_for**()
Devuelve un data frame con nombres y etiquetas

df %>% **look_for**("s")
Busca variables con "s" en su nombre o etiqueta

df %>% **look_for**(details = TRUE)
Devuelve detalles adicionales de cada variable

Etiquetas de valores

Al añadir etiquetas de valor a un vector numérico o de caracteres, su clase pasa a ser **haven_labelled**. A diferencia de un factor, los valores del vector no cambian y no es obligatorio asignar una etiqueta a cada valor.

MANIPULAR UN VECTOR

val_label(x, value) *o* **val_label**(df$v1, value)
Obtiene la etiqueta de un valor específico de un vector

val_label(x, value) <- "label"
Define/actualiza la etiqueta de un valor específico

val_label(x, value) <- NULL
Elimina la etiqueta de un valor específico

val_labels(x)
Obtiene todas las etiquetas de valor de un vector

val_labels(x) <- c(no = 0, yes = 1, maybe = 9)
Define/actualiza todas las etiquetas de valor de un vector

val_labels(x) <- NULL
Elimina todas las etiquetas de valor de un vector

labelled(c("F", "F", "M"), c(Female = "F", Male = "M"))
Crea un vector etiquetado

sort_val_labels(x, according_to = "values")
Ordena etiquetas de valor por valores (o etiquetas)

drop_unused_value_labels(x)
Elimina etiquetas de valores ausentes en los datos

MANIPULAR UN DATA.FRAME

df %>% **set_value_labels**(v1 = c(Yes = 1, No = 2), v2 = c(Male = "M", Female = "F"))
Define las etiquetas de valor de varias variables

df %>% **add_value_labels**(v1 = c(Unknown = 9))
Añade etiquetas de valor específicas a una variable (las etiquetas ya definidas no cambian)

df %>% **remove_value_labels**(v1 = 9)
Elimina etiquetas de valor específicas de una variable

df %>% **set_value_labels**(v1 = NULL)
Elimina todas las etiquetas de valor de una variable

df %>% **drop_unused_value_labels**()
Elimina etiquetas de valores ausentes en los datos

Valores faltantes

FALTANTES DEFINIDOS POR EL USUARIO (SPSS)

Sirven para indicar que algunos valores deben considerarse faltantes. Sin embargo, no se tratarán como NA mientras no se conviertan en NA propiamente dichos.

Al añadir valores faltantes a un vector numérico o de caracteres, su clase pasa a ser **haven_labelled_spss**.

Al importar un archivo SPSS, usa *user_na = TRUE* para conservar los faltantes definidos (si no, pasan a NA).

na_values(x)
Obtiene los valores faltantes individuales de un vector

na_values(x) <- c(8, 9, 10)
df %>% **set_na_values**(v1 = c(8, 9, 10))
Define/actualiza faltantes individuales (NULL = quitar)

na_range(x)
Obtiene el rango de valores faltantes de un vector

na_range(x) <- c(8, 10)
df %>% **set_na_range**(v1 = c(8, 10))
Define/actualiza un rango de faltantes (NULL = quitar)

user_na_to_na(x) *o* df %>% **user_na_to_na**()
Convierte en NA los faltantes definidos por el usuario

is_na(x)
TRUE si es NA o un valor faltante definido por el usuario

NA ETIQUETADOS (ESTILO STATA Y SAS)

Los faltantes "etiquetados" funcionan igual que los NA normales de R, pero guardan un byte adicional de información: una etiqueta, normalmente una letra ("a" a "z").

x <- c(1:5, tagged_na("a"), tagged_na("z"), NA)
tagged_na("a") genera un NA con etiqueta

is.na(x)
Los NA etiquetados funcionan igual que los NA normales

is_tagged_na(x)
Comprueba si es un NA etiquetado

na_tag(x)
Muestra las etiquetas de los NA etiquetados

format_tagged_na(x)
Convierte x en caracteres mostrando los NA etiquetados

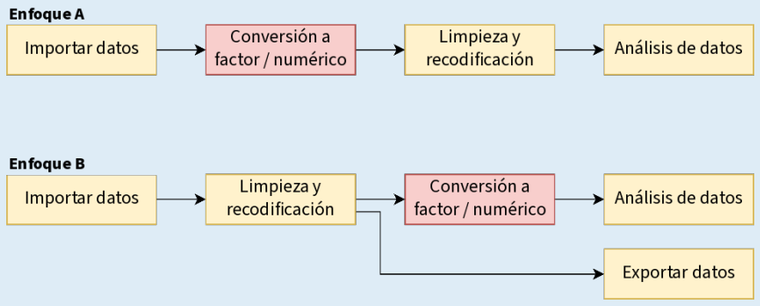
¿Cuándo usar etiquetas?

haven_labelled y *haven_labelled_spss*, clases del paquete **haven**, permiten añadir metadatos (etiquetas de variable, etiquetas de valor y valores faltantes estilo SPSS) a vectores o columnas de data frames, e importar correctamente esos metadatos desde SAS, Stata o SPSS.

Las funciones y métodos del paquete **labelled** están pensados para manipular fácilmente estos datos etiquetados.

Ten en cuenta que las **etiquetas de valor** no implican que los vectores sean categóricos o continuos. Por eso, no están pensadas para el análisis. Por ejemplo, antes de modelar o graficar, convierte los vectores con etiquetas de valor en factores o en vectores numéricos/de caracteres clásicos.

Se pueden considerar dos enfoques principales:



En el **enfoque A**, los vectores etiquetados se convierten en factores o vectores numéricos/de caracteres justo tras importar los datos, con **unlabelled**(), **to_factor**() o **unclass**(). Luego, la limpieza, recodificación y análisis usan los tipos de vectores clásicos de **R**.

En el **enfoque B**, los vectores etiquetados se conservan durante la limpieza y recodificación para preservar la codificación original, sobre todo si los datos se reexportarán. Las funciones de **labelled** ayudan a manejar las etiquetas de valor.

Pero, como en el enfoque A, habrá que convertirlos en factores o vectores numéricos clásicos antes del análisis, pues así deben codificarse las variables categóricas y continuas.

Conversión

DE VECTORES ETIQUETADOS

Si se eliminan todas las etiquetas de valor y los faltantes definidos por el usuario de un vector etiquetado, se quitará la clase *haven_labelled* y el vector se convertirá en un vector numérico o de caracteres básico. **Los valores del vector no cambian.**

remove_val_labels(x)
Elimina las etiquetas de valor de un vector.

remove_user_na(x)
Elimina los faltantes definidos por el usuario (*na_values* y *na_range*) de un vector

unclass(x)
Elimina la clase *haven_labelled*: el vector pasa a ser numérico o de caracteres clásico. Las etiquetas de valor y los faltantes definidos por el usuario siguen visibles como atributos.

Al convertir un vector etiquetado en un factor o en un vector de caracteres con las etiquetas de valor, ten en cuenta que los valores **originales del vector se convertirán.**

to_character(x)
Convierte en un vector de caracteres, reemplazando los valores por su etiqueta de valor

to_factor(x)
Convierte en un factor, reemplazando los valores por su etiqueta de valor

to_factor(x, levels = "prefixed")
Las etiquetas de valor llevarán como prefijo su valor original

to_factor(x, strict = TRUE)
Convierte en factor solo si todos los valores tienen etiqueta

df %>% **to_factor**()
Convierte en factores todos los vectores etiquetados, y solo ellos

df %>% **to_factor**(labelled_only = FALSE)
Convierte en factores todas las columnas (incluso no etiquetadas)

unlabelled(x)
df %>% **unlabelled**()
Los vectores etiquetados pasan a factores si todos sus valores tienen etiqueta; si no, se les quita la clase. Equivale a df %>% to_factor(labelled_only = T, strict = T, unclass = T)

to_factor(x, drop_unused_labels = TRUE)
df %>% **unlabelled**(drop_unused_labels = TRUE)
Quita etiquetas de valor no usadas antes de convertir en factores

A VECTORES ETIQUETADOS

to_labelled(f)
Convierte un factor en un vector etiquetado numérico.
Ten en cuenta que to_labelled(to_factor(x)) *y* x *no serán idénticos (se pierde la codificación original).*

to_labelled(df)
Si df se importó con **foreign** o se creó con **memisc**, sus metadatos (etiquetas de variable, de valor y faltantes definidos por el usuario) pasan al formato **labelled**.

Al añadir una etiqueta de valor o un faltante del usuario a un vector numérico o de caracteres, este se convierte en un vector etiquetado.
Los valores del vector no cambian.

Otras funciones

nolabel_to_na(x) *o* df %>% **nolabel_to_na**()
En vectores etiquetados, los valores sin etiqueta de valor se convertirán en NA

val_labels_to_na(x) *o* df %>% **val_labels_to_na**()
En vectores etiquetados, los valores con etiqueta de valor se convertirán en NA

df2 %>% **copy_labels_from**(df1)
Copia etiquetas de variable, etiquetas de valor y faltantes definidos por el usuario de df1 a df2 según los nombres de columna en común. Útil cuando los atributos se pierden tras manipular los datos.

recode(x, `2` = 1, `3` =2)
Aplica dplyr::recode() a un vector etiquetado.
Las etiquetas de valor no cambian.

recode(x, `2` = 1, .combine_value_labels = TRUE)
Esta opción combina las etiquetas de los valores originales fusionados para crear nuevas etiquetas de valor.
Se recomienda verificar que el resultado sea adecuado.

update_labelled(x) *o* df %>% **update_labelled**()
Si x/df se importó o creó con una versión antigua de **haven** o **labelled**, podrías obtener resultados inesperados.
update_labelled() actualiza todos los vectores etiquetados para que sean coherentes con la implementación actual.

CC BY SA Joseph Larmarange • Más información en **http://larmarange.github.io/labelled** • versión de labelled 2.5.0 • Actualizado: 2020-06 • Traducido por Anthony Romero-Cerdán • anthonyromero.netlify.app