

practica-2

guada

26/2/2020

Practica 2: Aprendiendo Tidyverse con dataset de victimas del terrorismo de estado

Actividades:

1. Subir a R el dataset de “Listado de víctimas del accionar represivo ilegal” elaborado por el Ministerio de Justicia y Derechos Humanos (<https://datos.gob.ar/dataset/justicia-registro-unificado-victimas-terrorismo-estado--ruvte->) que contiene un listado de las desapariciones y asesinatos ocurridos durante 1966 y 1983

(Recomendaciones: Como podemos ver el dataset tiene inconvenientes con los caracteres especiales, eso significa que tenemos que cambiarle el encoding.. el mismo se hace agregando dentro de la funcion `read.csv(path = , encoding = “Latin-1”)`)

2. Realizamos un `head()` para ver las variables
3. ¿Qué tipo de variables tiene la tabla? Escribirlo (numericas, caracteres, etc)
4. Carguemos la librería de tidyverse y conozcamos más sobre la información que brinda el dataset
 - a. Queremos saber, en principio, cuál es el país con mayor cantidad de desapariciones forzadas según la nacionalidad.
 - b. ¿Cuál es la provincia con mayor cantidad de desapariciones forzadas?
 - c. Cuántos años tendrían al día de hoy las personas registradas en el dataset cuya provincia de nacimiento sea “BUENOS AIRES” y que figure en el nuevo dataset sólo las columnas de “provincia_nacimiento”, “anio_nacimiento” y la “edad_actual”. b.a.¿Cuál es la edad promedio que tendrían al día de hoy?
5. ¿Qué pasa si queremos saber el género de la persona y no figura en el dataset? Vamos paso a paso.
 - a. Podemos empezar construirlo gracias al registro realizado por el gobierno de la Ciudad que tendremos que cargar en nuestro environment y que esta cargado en el github. (Acordemosnos de modificar el encoding, que en este caso es: “UTF-8”).

```
nombres <- read.csv("https://raw.githubusercontent.com/Guadag12/R4RRII/master/Clase%202/nombres.csv", en
```

- b. La idea es realizar un join y unir ambas tablas, sin embargo no tenemos una columna en nuestro dataset que unicamente sean nombres. Hay que construirla con la funcion “`cSplit()`” del paquete `splitstackshape`

```

#library(splitstackshape)
#data_terrorismo_arg <- cSplit(indt = data_terrorismo_arg, splitCols = 'apellido_paterno_nombres', sep = '_')

#llamamos nombre a la columna que nos interesa
#names(data_terrorismo_arg)[19] <- "nombre"

#vamos a trabajar con las columnas que nos interesan
#data_terrorismo_arg <- data_terrorismo_arg %>% select(anio_denuncia, tipificacion_ruute, id_unico_ruute,
#documentos, anio_nacimiento, provincia_nacimiento, pais_nacimiento, nacionalidad, embarazo,
#fecha_lugar_detencion_secuestro, fecha_lugar_asesinato_o_hallazgo_de_restos, fotografia,
#provincia_nacimiento_indec_id, pais_nacimiento_indec_id, apellido_paterno_nombres_1,
#nombre )

```

d. Realizamos el join y quitamos los duplicados!

e. ¿Podemos conocer el porcentaje de personas desaparecidas según el género?