

# Feminicidios en Colombia 2017 - Datos

Daniel Delgado

## Table of contents

1	Carga de los datos	1
2	Exploración e identificación de problemas	2
3	Limpieza de los datos	3
4	Guardado de los datos	4

## 1 Carga de los datos

En primer lugar, se deben cargar los paquetes necesarios para la importación, limpieza y organización de los datos. Para este caso, se utilizaron los paquetes `dplyr`, `tidyr` y `readr`.

```
library(dplyr)
library(tidyr)
library(readr)
```

Luego, importé los datos. Estos se encuentran en formato CSV y se encuentran en la carpeta `data`.

```
raw_femicides <- read_csv(
  "data/cases_2017.csv",
  show_col_types = FALSE
)
```

## 2 Exploración e identificación de problemas

Este conjunto de datos contiene información sobre casos de feminicidios ocurridos en Colombia durante el año 2017. Cuenta con un total de 149 observaciones y 15 variables. A continuación, se muestra una observación general de los datos.

```
glimpse(raw_femicides)
```

```
Rows: 149
Columns: 15
$ fecha                <date> 2017-01-03, 2017-01-03, 2017-01-05, ~
$ victima              <chr> "Diliana Bolaños Ijají", "Odali Bola~
$ edad                <chr> "24", "Sin información", "27", "28", ~
$ delito              <chr> "Feminicidio", "Feminicidio", "Femin~
$ modalidad           <chr> "Arma de fuego", "Arma de fuego", "0~
$ departamento        <chr> "Cauca", "Cauca", "Pasto", "Cauca", ~
$ municipio           <chr> "Balboa", "Balboa", "Ipiales", "Sant~
$ barrio              <chr> "Fátima", "Fátima", "La Pradera", "S~
$ relacion_victima     <chr> "Expareja", "Excuñado", "Pareja", "E~
$ `nombre agresor/presunto agresor` <chr> "Sin información", "Sin información"~
$ comentario          <chr> NA, NA, "El agresor es cobijado con ~
$ fuente              <chr> "http://www.elpais.com.co/judicial/c~
$ `fuente 2`          <chr> NA, NA, NA, NA, NA, NA, NA, NA, ~
$ longitud            <chr> "-77.216.747", "-77.216.747", NA, "-~
$ latitud             <chr> "2.041.023", "2.041.023", NA, "300.7~
```

A partir de lo anterior, noté que a excepción de la columna `fecha`, las demás columnas se importaron en formato carácter. Por lo que tendría sentido realizar los siguientes cambios en el tipo de dato de las variables:

- `edad`: entero
- `latitud`: numérico
- `longitud`: numérico

Además, en varias columnas, los valores faltantes se representan con el carácter `Sin información` y en otras con el valor `NA`. Por lo que para estandarizar los datos, sería conveniente reemplazar los valores `Sin información` por `NA`.

También noté, que los valores de las columnas `latitud` y `longitud` parecen haber sido mal digitados, por lo que es necesario realizar una transformación de los datos para que estos se encuentren en el formato correcto.

### 3 Limpieza de los datos

A continuación, se realiza la limpieza de los datos. En primer lugar, se reemplazan los valores Sin información por NA en todas las columnas.

```
femicides <- raw_femicides |>
  mutate(
    across(
      .cols = where(is.character),
      .fns = function(x) {
        if_else(tolower(x) == "sin información", NA, x)
      }
    )
  )
```

Posteriormente, se realiza la transformación de los datos de las columnas `latitud` y `longitud` para que estos se encuentren en el formato correcto.

```
fix_long <- function(longitud) {
  no_dots <- gsub("\\\\.", "", longitud)

  fixed_long <- paste0(
    substr(no_dots, 1, 3),
    ".",
    substr(no_dots, 4, nchar(no_dots))
  )
}

fix_lat <- function(latitud, departamento) {
  no_dots <- gsub("\\\\.", "", latitud)

  if_else(
    departamento == "Atlántico" | departamento == "Bolívar",
    fixed_lat <- paste0(
      substr(no_dots, 1, 2),
      ".",
      substr(no_dots, 3, nchar(no_dots))
    ),
    fixed_lat <- paste0(
      substr(no_dots, 1, 1),
      ".",
      substr(no_dots, 2, nchar(no_dots))
    )
  )
}
```

```

    )
  )
}

femicides <- femicides |>
  mutate(
    longitud = fix_long(longitud),
    latitud = fix_lat(latitud, departamento)
  )

```

Ahora continuaré con las modificaciones en el tipo de dato de las variables siguiendo lo anteriormente mencionado.

```

femicides <- femicides |>
  mutate(
    edad = as.integer(edad),
    latitud = as.numeric(latitud),
    longitud = as.numeric(longitud)
  )

```

También modificaré los nombres de algunas columnas para eliminar los espacios y caracteres especiales.

```

femicides <- femicides |>
  rename(
    presunto_agresor = `nombre agresor/presunto agresor`,
    fuente_2 = `fuente 2`,
  )

```

## 4 Guardado de los datos

Finalmente, se guardan los datos limpios en formato CSV en la carpeta `data`.

```

write_csv(femicides, "data/feminicidios.csv")

```