

Problem Set 3

Maria Jose Medellín (202122960), Juan Camilo Bejarano (202121927), Rafael Sánchez (202123676)

Versión de R: *R version 4.3.1 (2023-06-16 ucrt)*

Link al repositorio: <https://github.com/rafael-s-t/Problem-set-3.git>

Configuración inicial

```
## Set Up inicial
rm(list=ls())

##Cargar los paquetes y librerías necesarias
require(pacman)
p_load(tidyverse, rio, data.table, stringr, ggplot2)
```

Punto 1: Bucle

1.1. Lista de archivos en input

```
##Obtener la lista de archivos en input
rutas <- list.files("input" , recursive=T , full.names=T)
```

El objeto *rutas* almacena el vector de nombre de cada archivo dentro de la carpeta *input*. Cada archivo contiene la ruta de su ubicación.

1.2. Importar archivos

```
## Extraer las rutas
rutas_resto <- str_subset(string = rutas, pattern = "Resto - Ca")

## Cargar en lista
lista_resto <- vector("list", length = length(rutas_resto))

## Loop para cargar cada archivo
for (i in seq_along(rutas_resto)) {
  lista_resto[[i]] <- import(file = rutas_resto[i])
  lista_resto[[i]]$path <- rutas_resto[i]
  lista_resto[[i]]$year <- str_sub(rutas_resto[i], start = 7, end = 10)
}
```

A cada dataframe se añade una variable que permite identificar el año al que pertenece el archivo.

1.3. Combinar conjunto de datos

```
##Eliminar el elemento número 36 de la lista
lista_resto[[36]] <- NULL

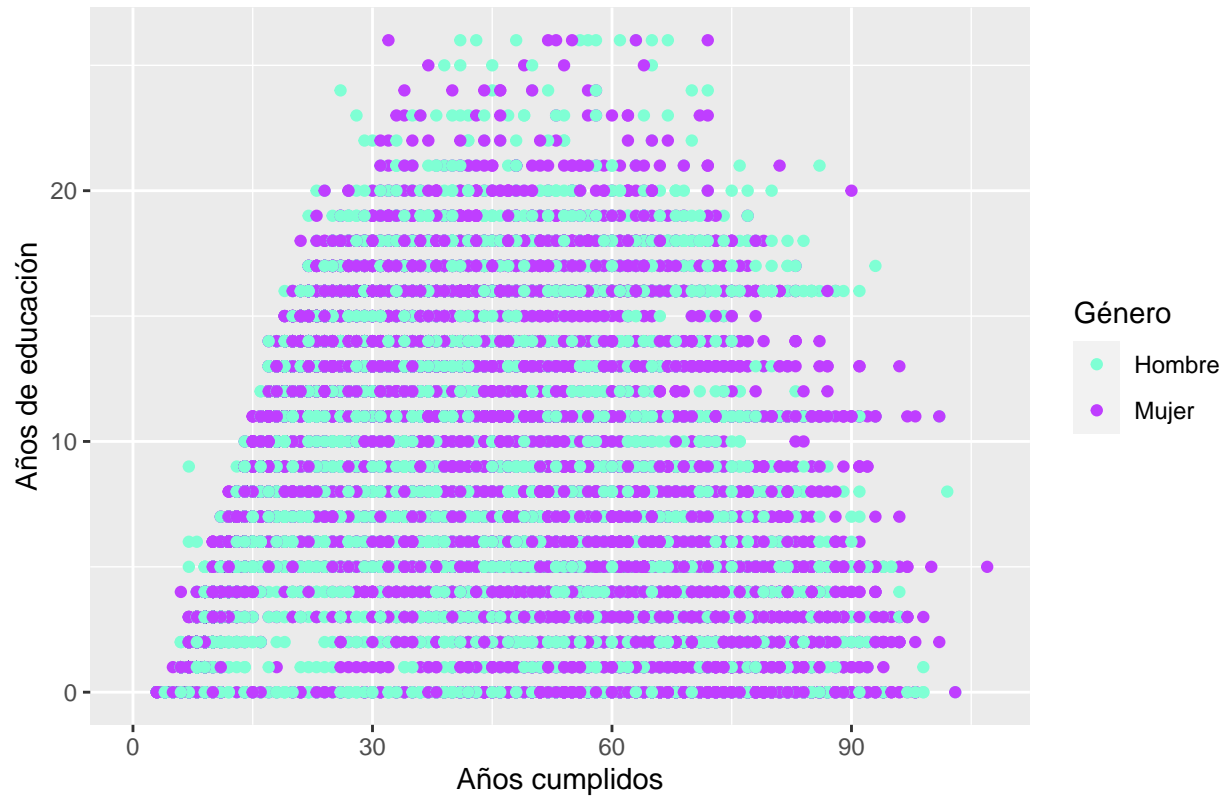
# Combinar la lista en un solo data frame
cg <- rbindlist(l=lista_resto , use.names=T , fill=T)
```

Nuevo data frame que contiene todos los anteriores es *cg*

Punto 2: Visualizaciones

Gráfica 1

Años cumplidos y años de educación, diferenciado por género



Gráfica 2

