

Problemset 3

Ximena Karaman y Luisa Perdomo

2023-12-06

1. Bucle

Para esta parte del taller, se presentan los códigos utilizados y como todos están en un mismo chunk, se presentan las tres partes de este punto como uno solo.

1.1 Lista de archivos en input

1.2 Importar archivos: “resto - cara”

1.3 Combinar conjunto de datos

```
setwd("/Users/luisaperdomo/Downloads/pset-3/input")
getwd()

## [1] "/Users/luisaperdomo/Downloads/pset-3/input"

nombres_archivos <- list.files(recursive = T , full.names = T)

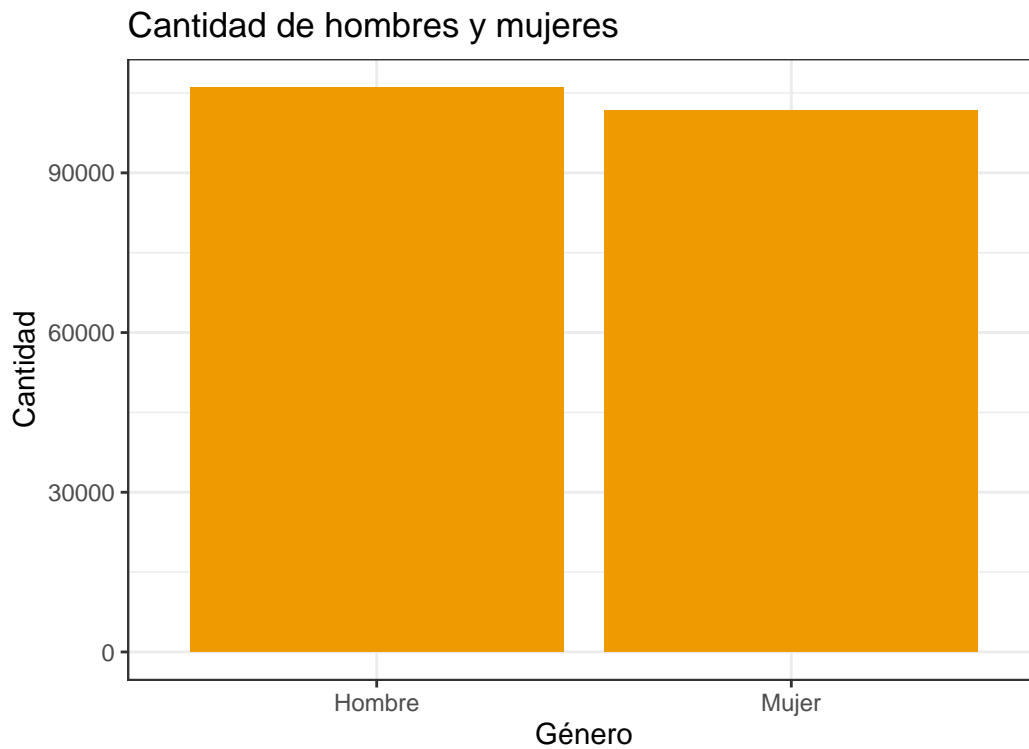
df <- import_list(str_subset(string = nombres_archivos , pattern = "Resto - Cara"))
for (i in 1:36){
  ruta_i <- str_subset(string = nombres_archivos , pattern = "Resto - Cara")[i]
  df[[i]]$ruta <- ruta_i
}

cg <- rbindlist(l=df[1:35], fill=TRUE)
```

Visualizaciones

Para esta parte, únicamente se presenta el output.

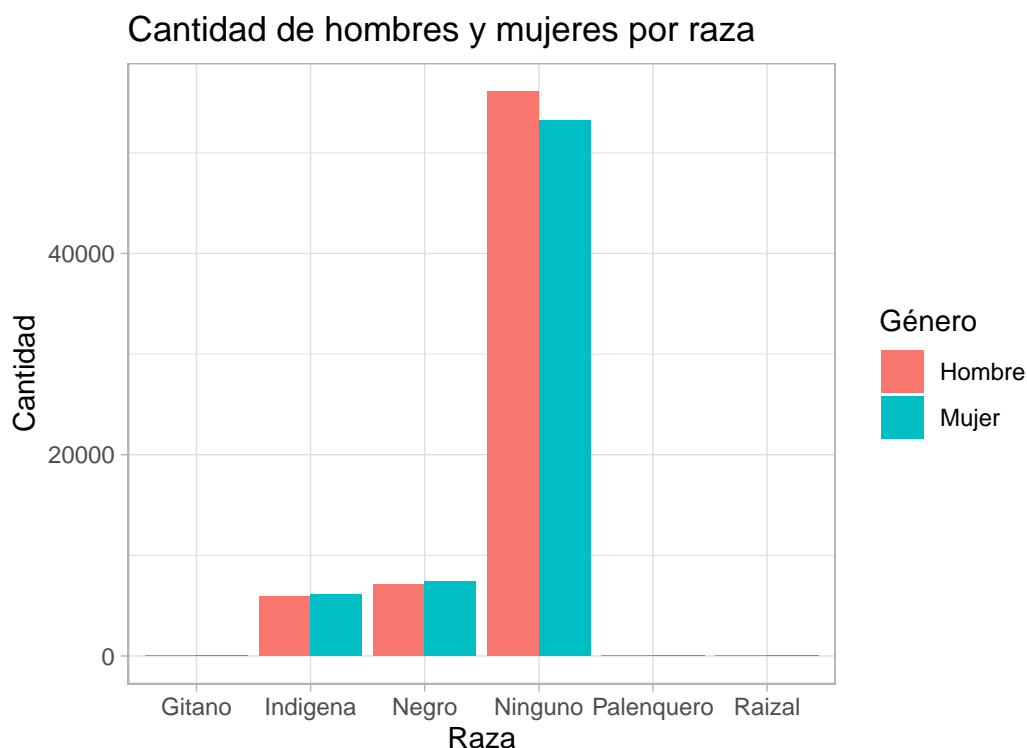
Visualización 1



Para esta primera visualización, fue necesario reemplazar los valores de la columna correspondiente a género de acuerdo al diccionario de la Gran Encuesta Integradora de Hogares, donde se establecía que todas aquellas celdas que tomaran el valor de 1 era igual a hombre y todas las que tomaran el valor de 2 era igual a mujer. Después de hacer el replace correspondiente, se utilizó el paquete ggplot2 para obtener el gráfico de barras correspondiente.

A manera de análisis, puede decirse que aunque la cantidad de hombres y mujeres encuestados en la GEIH es muy similar para ambos géneros, hay una mayoría de hombres y por lo tanto una minoría en mujeres.

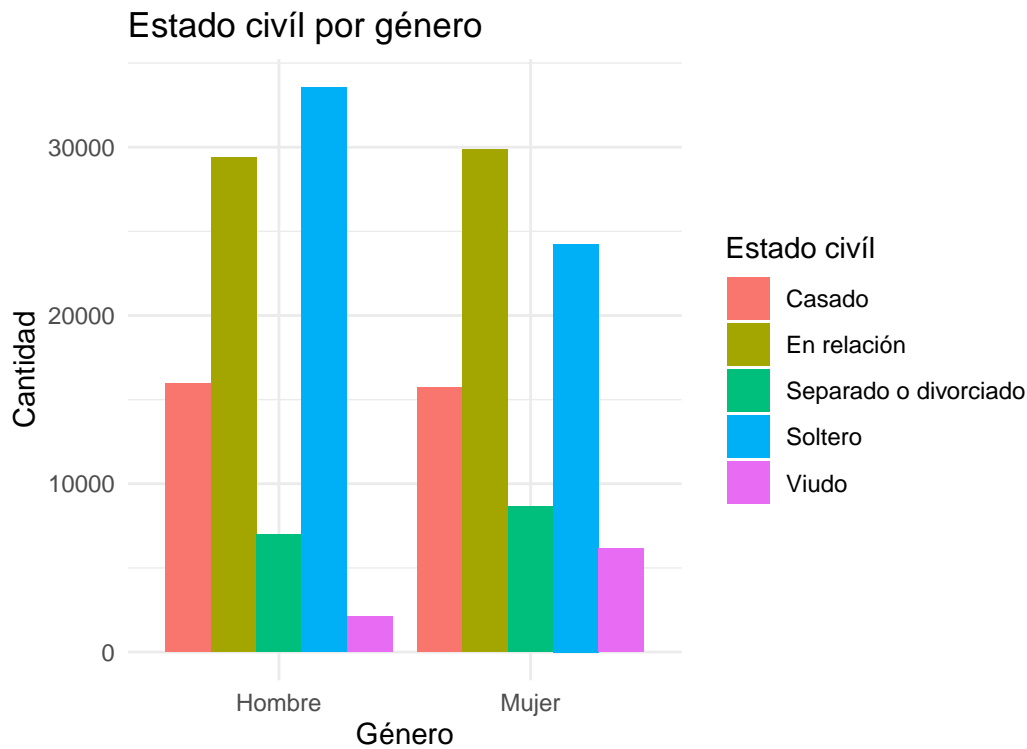
Visualización 2



Para esta segunda visualización, fue necesario reemplazar los valores de la columna correspondiente a raza de acuerdo al diccionario de la Gran Encuesta Integradora de Hogares, donde se establecía que todas aquellas celdas que tomaran el valor de 1 era igual a indígena, todas las que tomaran el valor de 2 era igual a gitano, todas las que tomaran el valor de 3 era igual a raizal, todas las que tomaran el valor de 4 era igual a palenquero, todas las que tomaran el valor de 5 era igual a negro y las que tomaran el valor igual a 6 era igual a ninguno, es decir que el encuestado no se siente identificado con las cinco opciones planteadas en cuanto a raza. Después de hacer el replace correspondiente, fue necesario filtrar todos los missing values y finalmente, se utilizó el paquete ggplot2 para obtener el gráfico de barras correspondiente.

A manera de análisis, es posible observar que la gran mayoría de encuestados, tanto hombres como mujeres, en la GEIH no se identifican con las opciones planteadas de raza. Sin embargo, aquellos que si puede verse que se identifican como negro o como indígena; esto de igual forma para ambos géneros.

Visualización 3



Para esta última visualización, fue necesario reemplazar los valores de la columna correspondiente a estado civil y de acuerdo al diccionario de la Gran Encuesta Integradora de Hogares, donde se establecía que todas aquellas celdas que tomaran el valor de 1 y 2 era igual a encontrarse en una relación, todas las que tomaran el valor de 3 era igual a estar casado, todas las que tomaran el valor de 4 era igual a estar separado o divorciado, todas las que tomaran el valor de 5 era igual a estar viudo, y todas las que tomaran el valor de 6 era igual a estar soltero. Después de hacer el replace correspondiente, fue necesario filtrar todos los missing values, para así utilizar el paquete ggplot2 y obtener el gráfico de barras correspondiente.

A manera de análisis, puede evidenciarse que para la sección de los hombres en lo que refiere se estado civil, la mayoría están solteros o en relación, y la minoría son viudos. Por el lado de las mujeres, el estado civil con mayor participación son en relación o soltera, y con minoría de viuda. Si bien ambos para ambos géneros las proporciones son similares, se observa que el comportamiento respecto al estado civil es muy parecido.