

# SOC4001 Procesamiento avanzado de bases de datos en R

## Tarea 3, respuestas

Ponderación: 12% de la nota final del curso Entrega: Desde el momento de entrega, los estudiantes tiene 1 exacta semana de plazo para completar esta tarea. Formato: Desarrollar esta tarea en un RScript, agregando comentarios cuando sea necesario.

- 1) Carga la base de datos “Chile” del paquete `carData` y crea un objeto que los contenga los datos. Llama tal objeto “datos\_chile”. Carga la librería `tidyverse` y ejecuta la siguientes operaciones usando las herramientas contenidas de `tidyverse`:

```
library("carData")
library("tidyverse")
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.1
## v tidyr   1.1.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
data("Chile")
datos_chile <- Chile
rm(Chile) # remueve "flotante"
```

```
datos_chile %>% glimpse()
```

```
## Rows: 2,700
## Columns: 8
## $ region      <fct> N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, ...
## $ population  <int> 175000, 175000, 175000, 175000, 175000, 175000, 175000, ...
## $ sex         <fct> M, M, F, F, F, F, M, F, F, M, M, M, F, F, M, M, F, M, ...
## $ age         <int> 65, 29, 38, 49, 23, 28, 26, 24, 41, 41, 64, 19, 27, 46, ...
## $ education   <fct> P, PS, P, P, S, P, PS, S, P, P, P, S, PS, S, PS, S, PS, ...
## $ income      <int> 35000, 7500, 15000, 35000, 35000, 7500, 35000, 15000, 15...
## $ statusquo   <dbl> 1.00820, -1.29617, 1.23072, -1.03163, -1.10496, -1.04685...
## $ vote        <fct> Y, N, Y, N, N, N, N, N, U, N, Y, U, Y, Y, NA, A, N, U, Y...
```

- 2) Añade a “datos\_chile” un variable llamada “year” con valor 1988 en todas las filas

```
datos_chile <- datos_chile %>% mutate(year = 1988)
```

- 3) Calcula el año de nacimiento de cada individuo. Añade a “datos\_chile” un variable llamada “birthyear” que contenga esta información

```
datos_chile <- datos_chile %>% mutate(birthyear = year - age)
```

- 4) Usando la función `case_when()` añade a “datos\_chile” un variable llamada “cohort73” que tome valor 1 si la persona tenía 15 años o más el año del golpe de estado (1973) y valor 0 si tenía menos de 15. Trata las observaciones que no cumplan ninguna de estas condiciones como valores perdidos.

```
datos_chile <- datos_chile %>% mutate(cohort73 = case_when(birthyear <= (1973 - 15) ~ 1,
  birthyear > (1973 - 15) ~ 0)
)
```

- 5) Usando la función `group_by()` añade a “datos\_chile” un variable llamada “income\_region\_educ” que contenga el promedio de la variable “income” por región y nivel educacional.

```
datos_chile <- datos_chile %>% group_by(region,education) %>%
  mutate(income_region_educ = mean(income, na.rm = T))
```

```
datos_chile %>% glimpse()
```

```
## Rows: 2,700
## Columns: 12
## Groups: region, education [19]
## $ region      <fct> N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, ...
## $ population  <int> 175000, 175000, 175000, 175000, 175000, 175000, ...
## $ sex         <fct> M, M, F, F, F, F, M, F, F, M, M, M, F, F, M, M, ...
## $ age         <int> 65, 29, 38, 49, 23, 28, 26, 24, 41, 41, 64, 19, ...
## $ education   <fct> P, PS, P, P, S, P, PS, S, P, P, P, S, PS, S, PS, ...
## $ income      <int> 35000, 7500, 15000, 35000, 35000, 7500, 35000, 1...
## $ statusquo   <dbl> 1.00820, -1.29617, 1.23072, -1.03163, -1.10496, ...
## $ vote        <fct> Y, N, Y, N, N, N, N, N, U, N, Y, U, Y, Y, NA, A, ...
## $ year        <dbl> 1988, 1988, 1988, 1988, 1988, 1988, 1988, 1988, ...
## $ birthyear   <dbl> 1923, 1959, 1950, 1939, 1965, 1960, 1962, 1964, ...
## $ cohort73    <dbl> 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, ...
## $ income_region_educ <dbl> 19397.81, 52896.83, 19397.81, 19397.81, 32345.13...
```