

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 3

Ponderación: 12% de la nota final del curso

Entrega: Desde el momento de entrega, los estudiantes tienen plazo hasta el 20 de Octubre a las 23:59pm para completar esta tarea. Formato: Desarrollar esta tarea en un RScript, agregando comentarios cuando sea necesario.

El código a continuación carga la Base de Datos Histórica Proyectos Adjudicados ANID (ex-conicyt) y extrae una selección de variables que son almacenados en el objeto `data_anid`.

```
library("tidyverse")
library("readr")

path <- url("https://raw.githubusercontent.com/ANID-GITHUB/Historico-de-Proyectos-Adjudicados/da63cab4f")

data_anid <- read_delim(path, delim = ";")

data_anid <- data_anid %>% rename(codigo_proyecto = CODIGO_PROYECTO, anno = ANO_FALLO, sexo = SEXO, area = AREA)
```

Descripción de los datos: La Agencia Nacional de Investigación y Desarrollo (ANID) cada año adjudica financiamiento para proyectos en Ciencia y Tecnología a través de sus diferentes concursos. La base de datos denominada “BDH_Proyectos” contiene la información disponible de proyectos adjudicados por la Agencia (antes del 2020, CONICYT) desde el año 1982 hasta el 2020, con fecha de corte al 31 de diciembre del 2020. Cada fila representa una iniciativa adjudicada. Los datos deben verse así:

```
## # A tibble: 6 x 5
##   codigo_proyecto anno sexo   area                                monto
##   <dbl> <dbl> <chr> <chr>                                <dbl>
## 1      1820005   1982 HOMBRE CIENCIAS NATURALES                300
## 2      1820006   1982 HOMBRE CIENCIAS MEDICAS Y DE LA SALUD    130
## 3      1820009   1982 HOMBRE CIENCIAS NATURALES                506
## 4      1820010   1982 HOMBRE HUMANIDADES                    335
## 5      1820015   1982 HOMBRE CIENCIAS NATURALES                260
## 6      1820043   1982 HOMBRE CIENCIAS AGRICOLAS                464
```

- 1) Usando los comandos `group_by()` y `summarise()` produce la siguiente tabla y asígnala al objeto `tabla_1`. El resultado debe verse así:
- 2) Carga la base de datos con el IPC anual y guárdala en un objeto llamado `datos_ipc`. Para los años con valores perdidos en la variable `datos_ipc$ipc`, usa la función `fill()` para asignales el valor correspondiente al año siguiente. Conserva sólo las variables `anno` e `ipc`. Los datos deben verse así:

```
datos_ipc <- read_csv(" ")
```

- 3) Usando algunos de los comandos `_join()` junta los datos en `tabla_1` y `datos_ipc` preservando toda la información disponible en `tabla_1`. El resultado debe verse así:
- 4) Crea la nueva variable `monto_precios2021` multiplicando las variables `monto` e `ipc`. Posteriormente remueve las variables `monto` e `ipc`. El resultado debe verse así:

- 5) Usando el comando `pivot_wider()` transforma los datos de la siguiente manera.
- 6) Usa la función `replace_na()` para reemplazar los valores perdidos en las variables `HOMBRE` y `MUJER` por ceros. El resultado debe verse así:
- 7) Crea una nueva variable llamada `dif_hombremujer` que mida la diferencia entre el monto asignado a hombres y mujeres = `HOMBRE - MUJER`. Posteriormente conserva sólo las variables `anno`, `area` y `dif_hombremujer`. El resultado debe verse así:
- 8) Usando el comando `pivot_wider()` modifica la tabla producida en (7) y produce la siguiente tabla:
- 9) Elige el valor correspondiente a una celda cualquiera y describe la información que comunica.