

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 5

Ponderación: 12% de la nota final del curso

Formato: Desarrollar esta tarea en un RScript, agregando comentarios cuando sea necesario.

Instrucciones:

Usa el siguiente código para cargar la base de datos sobre Covid-19 usados en clase y seleccionar sólo las variables numéricas más el identificador de continente.

```
library("tidyverse")
library("cowplot")
library("lubridate")
library("viridis")

path <- url("https://raw.githubusercontent.com/mebucca/dar_soc4001/master/slides/class_12/covid_data.csv")
# leer archivo csv
covid_data <- read_delim(path, delim=";")
covid_data <- covid_data %>% select(continent, total_cases_per_million, total_deaths_per_million, new_
```

Referencia: Hasell, J., Mathieu, E., Beltekian, D. et al. A cross-country database of COVID-19 testing. Sci Data 7, 345 (2020). <https://doi.org/10.1038/s41597-020-00688-8> y utilizad

Los datos deben verse así:

```
## Rows: 56,748
## Columns: 5
## $ continent      <chr> "Asia", "Asia", "Asia", "Asia", "Asia", "Asia~
## $ total_cases_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ total_deaths_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ new_cases_per_million  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ new_deaths_per_million <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

Ejercicio

1. Usando las funciones `nest()` y `map()` del paquete `purrr`, para cada continente crea un matriz de correlaciones entre todas las variables.
 - Nota: para crear la matriz de correlaciones la función `map()` debe tomar el siguiente argumento: `.f = ~ cor(., use = "pairwise.complete.obs")`. Guarda la matriz en una nueva columna llamada `corrs` y asigna el resultado a un nuevo objeto llamado `miscors`.

El objeto `miscors` debe verse así:

```
miscors <- covid_data %>%
  nest(-continent) %>%
  mutate(corrs = map(.x=data, .f= ~ cor(., use = "pairwise.complete.obs")))

miscors
```

```
## # A tibble: 6 x 3
##   continent      data      corrs
##   <chr>         <list>    <list>
## 1 Asia         <tibble [13,528 x 4]> <dbl [4 x 4]>
## 2 Europe       <tibble [14,828 x 4]> <dbl [4 x 4]>
## 3 Africa       <tibble [13,637 x 4]> <dbl [4 x 4]>
## 4 North America <tibble [9,116 x 4]>  <dbl [4 x 4]>
## 5 South America <tibble [3,404 x 4]>  <dbl [4 x 4]>
## 6 Oceania      <tibble [2,235 x 4]>  <dbl [4 x 4]>
```

y cada matriz de correlación contenida en `corrs` de ve así

```
miscors$corrs[[1]]
```

```
##               total_cases_per_million total_deaths_per_million
## total_cases_per_million             1.00000000             0.02057843
## total_deaths_per_million             0.02057843             1.00000000
## new_cases_per_million                0.16873694             0.06326542
## new_deaths_per_million               0.16440194            -0.13163684
##               new_cases_per_million new_deaths_per_million
## total_cases_per_million             0.16873694             0.16440194
## total_deaths_per_million             0.06326542            -0.13163684
## new_cases_per_million                1.00000000            -0.02377809
## new_deaths_per_million              -0.02377809             1.00000000
```

2. Continúa trabajando con el objeto `miscors`. Usando las funciones `nest()` y `map()` del paquete `purrr`, crea nueva columna llamada `mean_cor` que contenga el promedio de cada matriz de correlaciones almacenada en la columna `corrs`.

- Nota: para el promedio de las correlaciones la función `map()` debe tomar el siguiente argumento: `.f = ~mean(., na.rm=T)`. Asigna el resultado a un nuevo objeto llamado `miscors`. El nuevo objeto `miscors` debe verse así:

```
miscors <- miscors %>%
  mutate(mean_cor = map(.x=corrs, .f= ~mean(., na.rm=T) ))
miscors
```

```
## # A tibble: 6 x 4
##   continent      data      corrs      mean_cor
##   <chr>         <list>    <list>    <list>
## 1 Asia         <tibble [13,528 x 4]> <dbl [4 x 4]> <dbl [1]>
## 2 Europe       <tibble [14,828 x 4]> <dbl [4 x 4]> <dbl [1]>
## 3 Africa       <tibble [13,637 x 4]> <dbl [4 x 4]> <dbl [1]>
## 4 North America <tibble [9,116 x 4]>  <dbl [4 x 4]> <dbl [1]>
## 5 South America <tibble [3,404 x 4]>  <dbl [4 x 4]> <dbl [1]>
## 6 Oceania      <tibble [2,235 x 4]>  <dbl [4 x 4]> <dbl [1]>
```

3. Usando la función `unnest()` crea un nueva base de datos que contenga solo la correlación promedio `mean_cor` por continente. Asigna el resultado a un nuevo objeto llamado `miscors`. El nuevo objeto `miscors` debe verse así:

```
miscors <- miscors %>%
  select(continent,mean_cor) %>%
  unnest()
miscors
```

```
## # A tibble: 6 x 2
```

```
##   continent    mean_cor
##   <chr>        <dbl>
## 1 Asia         0.283
## 2 Europe       0.376
## 3 Africa       0.239
## 4 North America 0.380
## 5 South America 0.353
## 6 Oceania      0.327
```

4. En base a los datos almacenados en `miscors` crea el siguiente gráfico usando la geometría `geom_bar(stat = "identity")`:

```
miscors %>% ggplot(aes(x=continent, y=mean_cor)) + geom_bar(stat = "identity", colour="grey", fill="blue") +
  labs(x="Continente", y="Correlación promedio", title="Mi primer gráfico :)") + theme_bw()
```

