

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 4

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:

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

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=";")
```

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:

```
## spec_tbl_df [56,748 x 50] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ iso_code                : chr [1:56748] "AFG" "AFG" "AFG" "AFG" ...
## $ continent               : chr [1:56748] "Asia" "Asia" "Asia" "Asia" ...
## $ location                : chr [1:56748] "Afghanistan" "Afghanistan" "Afghanistan" "Afgha
## $ date                    : Date[1:56748], format: "2019-12-31" "2020-01-01" ...
## $ total_cases              : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ new_cases                : num [1:56748] 0 0 0 0 0 0 0 0 0 0 ...
## $ new_cases_smoothed      : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ total_deaths             : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths               : num [1:56748] 0 0 0 0 0 0 0 0 0 0 ...
## $ new_deaths_smoothed     : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ total_cases_per_million  : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ new_cases_per_million    : num [1:56748] 0 0 0 0 0 0 0 0 0 0 ...
## $ new_cases_smoothed_per_million : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ total_deaths_per_million : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ new_deaths_per_million  : num [1:56748] 0 0 0 0 0 0 0 0 0 0 ...
## $ new_deaths_smoothed_per_million : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ reproduction_rate       : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ icu_patients             : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ icu_patients_per_million : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
## $ hosp_patients            : num [1:56748] NA NA NA NA NA NA NA NA NA NA ...
```

```

## $ hosp_patients_per_million      : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_icu_admissions          : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_icu_admissions_per_million : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_hosp_admissions         : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ weekly_hosp_admissions_per_million: num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_tests                   : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests                     : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_tests_per_thousand       : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_per_thousand         : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_smoothed             : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ new_tests_smoothed_per_thousand : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ tests_per_case                 : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ positive_rate                  : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ tests_units                    : chr [1:56748] NA NA NA NA ...
## $ stringency_index               : num [1:56748] NA 0 0 0 0 0 0 0 0 0 ...
## $ population                     : num [1:56748] 38928341 38928341 38928341 38928341 38928341 ..
## $ population_density             : num [1:56748] 54.4 54.4 54.4 54.4 54.4 ...
## $ median_age                     : num [1:56748] 18.6 18.6 18.6 18.6 18.6 18.6 18.6 18.6 18.6 18.6 ...
## $ aged_65_older                 : num [1:56748] 2.58 2.58 2.58 2.58 2.58 ...
## $ aged_70_older                 : num [1:56748] 1.34 1.34 1.34 1.34 1.34 ...
## $ gdp_per_capita                 : num [1:56748] 1804 1804 1804 1804 1804 ...
## $ extreme_poverty                : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ cardiovasc_death_rate          : num [1:56748] 597 597 597 597 597 ...
## $ diabetes_prevalence            : num [1:56748] 9.59 9.59 9.59 9.59 9.59 9.59 9.59 9.59 9.59 9.59 ...
## $ female_smokers                  : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ male_smokers                    : num [1:56748] NA NA NA NA NA NA NA NA NA NA NA ...
## $ handwashing_facilities         : num [1:56748] 37.7 37.7 37.7 37.7 37.7 ...
## $ hospital_beds_per_thousand     : num [1:56748] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ life_expectancy                : num [1:56748] 64.8 64.8 64.8 64.8 64.8 ...
## $ human_development_index        : num [1:56748] 498 498 498 498 498 498 498 498 498 498 ...
## - attr(*, "spec")=
## .. cols(
## ..   iso_code = col_character(),
## ..   continent = col_character(),
## ..   location = col_character(),
## ..   date = col_date(format = ""),
## ..   total_cases = col_double(),
## ..   new_cases = col_double(),
## ..   new_cases_smoothed = col_double(),
## ..   total_deaths = col_double(),
## ..   new_deaths = col_double(),
## ..   new_deaths_smoothed = col_double(),
## ..   total_cases_per_million = col_double(),
## ..   new_cases_per_million = col_double(),
## ..   new_cases_smoothed_per_million = col_double(),
## ..   total_deaths_per_million = col_double(),
## ..   new_deaths_per_million = col_double(),
## ..   new_deaths_smoothed_per_million = col_double(),
## ..   reproduction_rate = col_double(),
## ..   icu_patients = col_double(),
## ..   icu_patients_per_million = col_double(),
## ..   hosp_patients = col_double(),
## ..   hosp_patients_per_million = col_double(),
## ..   weekly_icu_admissions = col_double(),

```

```
## .. weekly_icu_admissions_per_million = col_double(),
## .. weekly_hosp_admissions = col_double(),
## .. weekly_hosp_admissions_per_million = col_double(),
## .. total_tests = col_double(),
## .. new_tests = col_double(),
## .. total_tests_per_thousand = col_double(),
## .. new_tests_per_thousand = col_double(),
## .. new_tests_smoothed = col_double(),
## .. new_tests_smoothed_per_thousand = col_double(),
## .. tests_per_case = col_double(),
## .. positive_rate = col_double(),
## .. tests_units = col_character(),
## .. stringency_index = col_double(),
## .. population = col_double(),
## .. population_density = col_double(),
## .. median_age = col_double(),
## .. aged_65_older = col_double(),
## .. aged_70_older = col_double(),
## .. gdp_per_capita = col_double(),
## .. extreme_poverty = col_double(),
## .. cardiovasc_death_rate = col_double(),
## .. diabetes_prevalence = col_double(),
## .. female_smokers = col_double(),
## .. male_smokers = col_double(),
## .. handwashing_facilities = col_double(),
## .. hospital_beds_per_thousand = col_double(),
## .. life_expectancy = col_double(),
## .. human_development_index = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

Ejercicio

Produce un gráfico lo más parecido posible a la figura mostrada a continuación. Usa un `theme` y una paleta de colores de tu preferencia.

Pista: mi paleta de colores está definida por la siguiente línea de código: `scale_color_viridis(trans = "date", option = "plasma")` y mi theme es `dark_theme_gray()`, del paquete `library("ggdark")`. Ambos ejes están graficados en escala log10.

```
plot <- covid_data %>% filter(continent=="Europe") %>%
  ggplot(aes(x=new_cases_smoothed_per_million, y=new_deaths_smoothed_per_million, colour=date)) +
  geom_point(alpha=0.1) +
  scale_x_log10() + scale_y_log10() +
  dark_theme_gray() +
  scale_color_viridis(trans = "date", option = "plasma") +
  labs(x="New cases smoothed per million", y="New deaths smoothed per million", title="Covid-19 in Europe")
print(plot)
```

