

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 5

Ponderación: 12% de la nota final del curso

Entrega: Desde el momento de entrega, los estudiantes tienen plazo hasta el domingo 29 de Noviembre a las 23:59pm para completar esta tarea.

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

- 1) Carga la base de datos sobre Covid-19 usados en clase: link

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: 50
## $ iso_code          <chr> "AFG", "AFG", "AFG", "AFG", "AFG...
## $ continent         <chr> "Asia", "Asia", "Asia", "Asia", ...
## $ location          <chr> "Afghanistan", "Afghanistan", "A...
## $ date              <date> 2019-12-31, 2020-01-01, 2020-01-...
## $ total_cases        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_cases          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ new_cases_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, ...
## $ total_deaths       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_deaths         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ new_deaths_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, ...
## $ total_cases_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_cases_per_million <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ new_cases_smoothed_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, ...
## $ total_deaths_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_deaths_per_million <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ new_deaths_smoothed_per_million <dbl> NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, ...
## $ reproduction_rate <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ icu_patients       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ icu_patients_per_million <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ hosp_patients      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ hosp_patients_per_million <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ weekly_icu_admissions <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ weekly_icu_admissions_per_million <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ weekly_hosp_admissions <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ weekly_hosp_admissions_per_million <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ total_tests        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_tests          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ total_tests_per_thousand <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_tests_per_thousand <lgl> NA, NA, NA, NA, NA, NA, NA, NA, ...
```

```
## $ new_tests_smoothed      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ new_tests_smoothed_per_thousand <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ tests_per_case          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ positive_rate           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ tests_units             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ stringency_index        <dbl> NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ population              <dbl> 38928341, 38928341, 38928341, 38...
## $ population_density      <dbl> 54.422, 54.422, 54.422, 54.422, ...
## $ median_age              <dbl> 18.6, 18.6, 18.6, 18.6, 18.6, 18...
## $ aged_65_older           <dbl> 2.581, 2.581, 2.581, 2.581, 2.58...
## $ aged_70_older           <dbl> 1.337, 1.337, 1.337, 1.337, 1.33...
## $ gdp_per_capita           <dbl> 1803.987, 1803.987, 1803.987, 18...
## $ extreme_poverty         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ cardiovasc_death_rate    <dbl> 597.029, 597.029, 597.029, 597.0...
## $ diabetes_prevalence      <dbl> 9.59, 9.59, 9.59, 9.59, 9.59, 9...
## $ female_smokers            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ male_smokers              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ handwashing_facilities   <dbl> 37.746, 37.746, 37.746, 37.746, ...
## $ hospital_beds_per_thousand <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0...
## $ life_expectancy          <dbl> 64.83, 64.83, 64.83, 64.83, 64.8...
## $ human_development_index  <dbl> 498, 498, 498, 498, 498, 498, 49...
```

- 2) 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")`.

```
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)
```

