

Estudi dels contagis a Europa durant la pandèmia de la Covid-19 i la influència de la vacuna

Visualització de dades - PAC 3

Héctor Gutiérrez Muñoz

17 d'abril, 2021

Sumari

1	Càrrega del fitxer	1
2	Primera exploració	1
3	Preprocessament	2
4	Fitxer de sortida	4

1 Càrrega del fitxer

En primer lloc, carreguem els fitxers CSV.

```
covid_data <- read.csv(  
  "https://covid.ourworldindata.org/data/owid-covid-data.csv",  
  na.strings = "")  
  
covid_data <- covid_data[, c("iso_code", "continent", "location", "date",  
  "new_cases_smoothed", "new_deaths_smoothed",  
  "population")]  
  
vaccine_data <- read.csv(  
  "https://opendata.ecdc.europa.eu/covid19/vaccine_tracker/csv/data.csv",  
  na.strings = "",  
  fileEncoding = "UTF-8-BOM"  
)
```

2 Primera exploració

Veiem uns exemples del primer dataset:

```
kable(
  head(covid_data),
  format = 'latex',
  booktabs = TRUE
) %>%
kableExtra::kable_styling(latex_options = c("scale_down", "hold_position"))
```

iso_code	continent	location	date	new_cases_smoothed	new_deaths_smoothed	population
AFG	Asia	Afghanistan	2020-02-24	NA	NA	38928341
AFG	Asia	Afghanistan	2020-02-25	NA	NA	38928341
AFG	Asia	Afghanistan	2020-02-26	NA	NA	38928341
AFG	Asia	Afghanistan	2020-02-27	NA	NA	38928341
AFG	Asia	Afghanistan	2020-02-28	NA	NA	38928341
AFG	Asia	Afghanistan	2020-02-29	0.143	0	38928341

També del segon:

```
kable(
  head(vaccine_data),
  format = 'latex',
  booktabs = TRUE
) %>%
kableExtra::kable_styling(latex_options = c("scale_down", "hold_position"))
```

YearWeekISO	FirstDose	FirstDoseRefused	SecondDose	UnknownDose	NumberDosesReceived	Region	Population	ReportingCountry	TargetGroup	Vaccine	Denominator
2020-W53	0	0	0	0	0	AT	8901064	AT	1_Age60+	MOD	2259262
2020-W53	0	0	0	0	0	AT	8901064	AT	1_Age60+	JANSS	2259262
2020-W53	0	0	0	0	0	AT	8901064	AT	1_Age60+	AZ	2259262
2020-W53	2659	0	0	0	61425	AT	8901064	AT	1_Age60+	COM	2259262
2020-W53	0	0	0	0	0	AT	8901064	AT	1_Age<60	MOD	5099181
2020-W53	0	0	0	0	0	AT	8901064	AT	1_Age<60	JANSS	5099181

3 Preprocessament

El dataset de vacunes permet desagregar per grup de població i tipus de vacuna. Aquí només ens interessa la informació de vacunació total.

```
# Només ens interessen les files amb grup de població "ALL"

vaccine_data <- vaccine_data[vaccine_data$TargetGroup == "ALL",]

# Agrupem per data i país i eliminem la columna del tipus de vacuna
# També eliminem la resta de columnes, que tampoc no ens interessen

vaccine_data <- vaccine_data %>%
  group_by(YearWeekISO, Region) %>%
  summarise(FirstDose_sum = sum(FirstDose),
            SecondDose_sum = sum(SecondDose),
            UnknownDose_sum = sum(UnknownDose))
```

'summarise()' has grouped output by 'YearWeekISO'. You can override using the '.groups' argument.

```
vaccine_data$num_vacunes <-
  vaccine_data$FirstDose_sum +
  vaccine_data$SecondDose_sum +
  vaccine_data$UnknownDose_sum

vaccine_data <- vaccine_data[,c("YearWeekISO", "Region", "num_vacunes")]
```

Com que el dataset també té dades de regions dels països que a nosaltres no ens interessin, eliminem aquestes files del dataset. Notar que només els països tenen un codi de dues lletres.

```
vaccine_data <- vaccine_data[nchar(vaccine_data$Region) == 2, ]
```

Conversió de la data del dataset de vacunes, hi ha un offset en les setmanes de 2020. Com que aquest dataset només té les dades informades per setmanes, omplim també la resta de dies (tots els valors d'una mateixa setmana seran similars):

```
vaccine_data$YearWeekISO[vaccine_data$YearWeekISO == "2020-W52"] <- "2020-W51"
vaccine_data$YearWeekISO[vaccine_data$YearWeekISO == "2020-W53"] <- "2020-W52"

vaccine_data$dat_yyyymmdd <- strptime(
  strptime(paste(vaccine_data$YearWeekISO, 7), format = "%Y-W%U%u"),
  format = "%Y/%m/%d"
)

vaccine_week <- list()

for (i in 1:6) {
  aux <- vaccine_data
  aux$dat_yyyymmdd <- strptime(
    as.Date(aux$dat_yyyymmdd, format = "%Y/%m/%d") - i, format = "%Y/%m/%d"
  )
  vaccine_week[[i]] <- aux
}

vaccine_week[[7]] <- vaccine_data

vaccine_data <- do.call(rbind, vaccine_week[1:7])
```

Dels dos datasets, hem de triar només els països de la UE, i deixar-los amb el mateix format.

```
covid_data <- covid_data[covid_data$iso_code %in%
  c("AUT", "BEL", "BGR", "HRV", "CYP", "CZE", "DNK",
    "EST", "FIN", "FRA", "DEU", "GRC", "HUN", "IRL",
    "ITA", "LVA", "LTU", "LUX", "MLT", "NLD", "POL",
    "PRT", "ROU", "SVK", "SVN", "ESP", "SWE", "NOR",
    "ISL", "LIE"), ]

# EL no és un codi de país, es refereix a Grècia (GR)
vaccine_data[vaccine_data$Region == "EL",]$Region <- "GR"

vaccine_data$Region <- countrycode(vaccine_data$Region, "iso2c", "iso3c")
```

Comprovem que hem deixat els mateixos països

```
setdiff(unique(covid_data$iso_code), unique(vaccine_data$Region))
```

```
## character(0)
```

```
setdiff(unique(vaccine_data$Region), unique(covid_data$iso_code))
```

```
## character(0)
```

Així, ja podem fer el join:

```
# Mateix format en les dates
```

```
covid_data$date <- strptime(  
  strptime(covid_data$date, format = "%Y-%m-%d"),  
  format = "%Y/%m/%d"  
)
```

```
head(vaccine_data)
```

```
## # A tibble: 6 x 4  
## # Groups:   YearWeekISO [1]  
##   YearWeekISO Region num_vacunes dat_yyyymmdd  
##   <chr>         <chr>         <int> <chr>  
## 1 2020-W51      BGR              0 2020/12/19  
## 2 2020-W51      CYP              0 2020/12/19  
## 3 2020-W51      CZE            1252 2020/12/19  
## 4 2020-W51      DEU            23998 2020/12/19  
## 5 2020-W51      EST              3 2020/12/19  
## 6 2020-W51      GRC             240 2020/12/19
```

```
merged <- merge(covid_data, vaccine_data,  
  by.x = c("date", "iso_code"),  
  by.y = c("dat_yyyymmdd", "Region"),  
  all.x = TRUE)  
  
final_df <- merged[, c("date", "iso_code",  
  "new_cases_smoothed", "new_deaths_smoothed",  
  "population", "num_vacunes")]
```

4 Fitxer de sortida

Finalment, les dades preprocessades es desaran a un fitxer. Només en treurem les dades de les dues últimes setmanes ja que canvien molt i desvirtuen la visualització

```
final_df <- final_df[final_df$date <= '2021/05/02', ]  
  
kable(  
  head(final_df[!is.na(final_df$num_vacunes), ]),  
  format = 'latex',
```

```

    booktabs = TRUE
) %>%
kableExtra::kable_styling(latex_options = c("scale_down", "hold_position"))

```

	date	iso_code	new_cases_smoothed	new_deaths_smoothed	population	num_vacunes
8989	2020/12/14	BGR	2479.857	118.286	6948445	0
8990	2020/12/14	CYP	369.000	2.714	875899	0
8991	2020/12/14	CZE	4892.286	101.000	10708982	1252
8992	2020/12/14	DEU	22465.000	457.143	83783945	23998
8995	2020/12/14	EST	451.286	2.857	1326539	3
8996	2020/12/14	FIN	461.286	5.286	5540718	24

```

write.csv(final_df, "covid19data.csv", row.names = FALSE, na = "", quote = FALSE)

```