UNIVERSITÀ DEGLI STUDI DI BRESCIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Corso di Laurea Magistrale
in Ingegneria Informatica

"Current status of Covid-19 cases in Europe and the influence of the vaccination campaign"

Health Information Systems course project

Marcello Manenti

Anno Accademico 2020/2021

INDEX

1. INTRODUCTION

The current situation of the pandemic is looking way better than the same period last year: there are less restrictions, less cases and less ICU beds occupied. The hospitals and healthcare organizations are less stressed, and the population is slowly getting back to a normal way of living.

But which factor was introduced, that allowed the situation to get better? Since the end of December 2020, in most of the EU countries, an intense **vaccination campaign** has started.

The main aim of the approved vaccines, is to reduce the severity of the symptoms, relieving the pressure on the healthcare sector and allowing people to live a normal life again. The purpose of reducing the spread of the virus is secondary because, although the vaccination reduces the possibility of infection, it doesn't avoid it completely.

In this project, I've used two datasets containing data about two Covid-19 main topics, cases, and vaccines, to try and find a correlation between the evolution of the cases (and deaths) of Covid-19 and the rising number of fully vaccinated people.

The data was retrieved from the European Centre for Disease Prevention and Control website. Here are the links for the two datasets:

- *"Data on the daily number of new reported COVID-19 cases and deaths by EU/EEA country"*

  https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country

- *"Data on COVID-19 vaccination in the EU/EEA"*

  https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea

2. DATA DESCRIPTION

The data I imported is updated to November 5th 2021, using the *read.csv(…)* function.

```
> head(cases_full)
     dateRep day month year cases deaths countriesAndTerritories geoId countryterritoryCode popData2020 continentExp
1 05/11/2021   5    11 2021  8877     22                 Austria    AT                  AUT     8901064       Europe
2 04/11/2021   4    11 2021  6494     14                 Austria    AT                  AUT     8901064       Europe
3 03/11/2021   3    11 2021  4962     10                 Austria    AT                  AUT     8901064       Europe
4 02/11/2021   2    11 2021  4972     10                 Austria    AT                  AUT     8901064       Europe
5 01/11/2021   1    11 2021  5741      8                 Austria    AT                  AUT     8901064       Europe
6 31/10/2021  31    10 2021  6014     10                 Austria    AT                  AUT     8901064       Europe
> head(vaccine_full)
  YearWeekISO FirstDose FirstDoseRefused SecondDose UnknownDose NumberDosesReceived Region Population ReportingCountry TargetGroup Vaccine Denominator
1    2020-W53         0               NA          0           0                   0     AT    8901064               AT         ALL      AZ     7358443
2    2020-W53         0               NA          0           0                   0     AT    8901064               AT         ALL     MOD     7358443
3    2020-W53         0               NA          0           0                   0     AT    8901064               AT         ALL   JANSS     7358443
4    2020-W53      5149               NA          0           0               61425     AT    8901064               AT         ALL     COM     7358443
5    2020-W53         0               NA          0           0                   0     AT    8901064               AT         ALL     UNK     7358443
6    2020-W53         0               NA          0           0                   0     AT    8901064               AT      Age0_4     UNK          NA
```

Dimension of data is:

- Cases → 7 500 observations of 11 variables
- Vaccines → 115 900 observations of 12 variables

2.1 DATA DICTIONARY

2.1.1 Cases dataset

VARIABLE	DEFINITION	TYPE
dateRep	Date of reporting “dd/mm/yyyy”	String
day		unit8
month		unit8
year		unit8
cases	Number of newly reported cases	int64
Deaths	Number of newly reported deaths	int64
countriesAndTerritories	Name of the country or territory	String
geoId	2-letter code	String
countryterritoryCode	3-letter ISO code	String

popData2020	Eurostat 2020 data (population)	int64
continentExp	Name of the continent reporting	String

2.1.2 Vaccines dataset

VARIABLE	DEFINITION	TYPE
YearWeekISO	Date when the vaccine was received/administered. Only weeks are allowed (e.g., “2021-W01”)	yyyy-Www
ReportingCountry	ISO 3166-1-alpha-2 code of reporting country	Two letter code
Denominator	Population denominators for target groups (total population obtained from Eurostat/UN).	Numeric
NumberDosesReceived	Number of vaccine doses distributed by the manufacturers to the country during the reporting week	Numeric
FirstDose	Number of first dose vaccine administered to individuals during the reporting week	Numeric
FirstDoseRefused	Number of individuals refusing the first vaccine dose	Numeric
SecondDose	Number of second dose vaccine administered to individuals during the reporting week	Numeric
UnknownDose	Number of doses administered during the reporting week where the type of dose was not specified (i.e. it is not known whether it was a first or second dose)	Numeric
Region	As a minimum data should be reported at national level (Region = country code)	Country/NUTS1 or 2/GAUL1/Country specific
TargetGroup	Target group for vaccination. As a minimum the following should be reported:	ALL = Overall adults (18+) Age<18 = Overall adolescents and children

	“ALL” for the overall figures “HCW” for healthcare workers Age-groups (preferably using the detailed age-groups)	HCW = Healthcare workers LTCF = Residents in long term care facilities Age0_4 = 0-4 years old Age5_9 = 5_9 years old Age10_14 = 10_14 years old Age15_17 = 15_17 years old Age18_24 = 18_24 years old Age25_49 = 25_49 years old Age50_59 = 50_59 years old Age60_69 = 60_69 years old Age70_79 = 70_79 years old Age80+ = 80 years and over AgeUnk = Unknown age 1_Age<60 = adults below 60 years of age 1_age60+ = adults 60 years and over
Vaccine	Name of vaccine. Additional vaccines will be added on approval or as requested.	AZ = Vaxzevria – AstraZeneca BECNBG (previously CN) = Inactivated – Beijing CNBG BHACOV = Covaxin – Bharat COM = Comirnaty – Pfizer/BioNTech JANSS = Ad26.COV 2.5 – Janssen HAYATVAC = Hayat VAC MOD = mRNA-1273 – Moderna QAZVAQ = QazCovid-In SIICOV = Covishield – SII SIN = CoronaVac – Sinovac SPU = Sputnik V - Gamaleya SRCVB = EpiVacCorona – SRCVB WUCNBG = Inactivated – Wuhan CNBG UNK = UNKNOWN ZFUZ = Sino-Uzbek - ZF-UZVAC
Population	Age-specific population for the country	Numeric

3. DATA CLEANING

The R libraries and packages used in this project are:

- pracma
- dplyr
- ggplot2
- gridExtra
- ggmap
- ggrepel

3.1 CASES DATASET

The first thing I saw in the cases dataset, is the *continentExp* column: considering the data is all European, every record of this variable is “Europe”. I proved this by filtering the dataframe with the *continentExp!=”Europe”* condition, resulting in 0 rows, so I dropped the column.

The geographic data, identifying different EU countries, is contained in three columns:

- *countriesAndTerritories*
- *geoId*
- *countryterritoryCode*

For the purposes of the analysis I want to apply, I could delete two of the columns, but I chose to only delete *geoId* to keep a column of extended country names (*countriesAndTerritories*) and a column of 3 digit code (*countryterritoryCode*).

Concerning time information, the dataset contains the date both in the string format in the *dateRep* column, and in the integer format, in the three columns *day, month* and *year*.

Considering the analysis I want to apply, I can delete the three separate columns and only keep the *dateRep* column, converting it to a date format with the *as.Date(…)* function.

3.1.1 Preparing the data for the comparison

Before deleting the *day, month,* and *year* columns, I need to apply a few steps to compute the week number for every day in the dataset, to facilitate the comparison with the vaccine dataset, which contains data on a weekly basis.

Looking at the format adopted by the vaccine dataset, i.e., the ISO 8601 definition with the yyyy-Www format and the week starting on Monday, I need to calculate the number of the week relating to each date present in the dataset.

To start, I computed the number of each day, storing it in a temporary vector, called *ndays*. To do it, I had to calculate the number of days passed at the start of every month, by creating a for loop to manage the various cases. The vector *trentag* contains the month number of the months having 30 days.

```
ndays <- vector("numeric", length = nrow(cases_c))
trentag <- c(4,6,9,11)
month_days <- vector("numeric", length = 11)

for(i in 1:11){
  sum <- 0
  for(j in 1:i){
    if(j==2){
      sum <- sum + 28
    } else if(j%in%c(4,6,9,11)){
      sum <- sum + 30
    } else {
      sum <- sum + 31
    }
  }
  month_days[i] <- sum
}
```

Then, I computed the number of every date contained in the dataframe, by doing the sum of the number of days of previous months to the current day of the month.

The days number go backwards from the 309^{th} day (November 5^{th}) to the 60^{th} (March 1^{st}, day of the start of data collection), and then it repeats itself for every country.

Considering that the first week of 2021 starts on 04/01/2021, I can compute the week numbers by doing:

$$\left(\frac{\text{ndays} - 4}{7}\right) + 1$$

So, to create the same format of the vaccine dataset (yyyy-Www), I just put the computed numbers in a string with the same format. Finally, I sum up the data for every week.

```
> head(cases_weekly_tot)
  countriesAndTerritories YearWeekISO countryterritoryCode  cases deaths popData2020
1                 Austria     2021-W09                  AUT 469539   8640     8901064
2                 Austria     2021-W10                  AUT  18468    152     8901064
3                 Austria     2021-W11                  AUT  20737    184     8901064
4                 Austria     2021-W12                  AUT  22100    173     8901064
5                 Austria     2021-W13                  AUT  21652    199     8901064
6                 Austria     2021-W14                  AUT  19309    217     8901064
```

3.2 VACCINES DATASET

At a first glance, the *Region* column and the *ReportingCountry* column seemed equal, but in the *Region* column there are more specific codes for single regions of every country.

For my analysis, the data concerning single regions is not necessary, so I dropped all the rows containing regional data to avoid duplicates.

Furthermore, the first dose refusal data contained in the *FirstDoseRefused* column is of little value to my analysis. After analyzing the 49 894 values column, 45 771 values were found to be NA, i.e., more of the 91,7% of the values are unavailable. I deleted the column.

Then, I checked the types of vaccine present in the dataset:

```
> unique(vaccine_full$vaccine)
[1] "AZ"       "MOD"      "JANSS"   "COM"      "UNK"      "SPU"      "BECNBG"
```

AZ	Double dose
MOD	Double dose
JANSS	Single dose
COM	Double dose
UNK	Undefined
SPU	Double dose
BECNBG	Double dose

Given the single dose nature of the Janssen vaccine, I thought it was better to analyze it differently from other vaccines, so I created a separated dataframe for the single dose vaccine, to be able to identify these numbers as single, fully immunizing doses.

The age groups are very confusing, not all the countries include the same ranges, and many groups are overlapping, distorting the count.

```
> unique(vaccine_c$TargetGroup)
 [1] "ALL"       "Age0_4"    "Age10_14" "Age15_17" "Age18_24" "Age25_49" "Age50_59" "Age5_9"    "Age60_69"
[10] "Age70_79" "Age80+"    "Age<18"    "AgeUNK"    "HCW"       "LTCF"      "1_Age60+" "1_Age<60"
```

I have detected four macro groups that can be reproduced for all the data:

- < 18
- 18 – 49
- 50 – 79
- 80 +

```
> vaccine_age[1:14,]
   ReportingCountry TargetGroup FirstDose SecondDose
1                AT      Age<18    252822     217851
2                AT      Age0_4       286         95
3                AT    Age10_14     96660      81435
4                AT    Age15_17    153998     135527
5                AT    Age18_24    461259     368374
6                AT    Age25_49   2098804    1787773
7                AT      Age5_9      1878        794
8                AT    Age50_59   1100298     992117
9                AT    Age60_69    891930     839253
10               AT    Age70_79    649461     620306
11               AT      Age80+    485912     462652
12               AT         ALL   5687664    5070475
13               BE      Age<18    651094     613286
14               BE    Age18_24    743846     661047
```

```
> vaccine_age_c[1:6,]
  ReportingCountry_age TargetGroup FirstDose_age SecondDose_age
1                   AT      Age<18        252822         217851
2                   AT    Age18_49       2560063        2156147
3                   AT    Age50_79       2641689        2451676
4                   AT      Age80+        485912         462652
5                   BE      Age<18        651094         613286
6                   BE    Age18_49       3856797        3530097
```

Note: data from Germany (DE), Liechtenstein (LI) and Netherlands (NL) doesn't contain age group division, so it wasn't included in this dataframe

4. EXPLORATORY DATA ANALYSIS

The exploratory data analysis in this project, was mainly done using the *ggplot2* library.

4.1 CASES DATASET

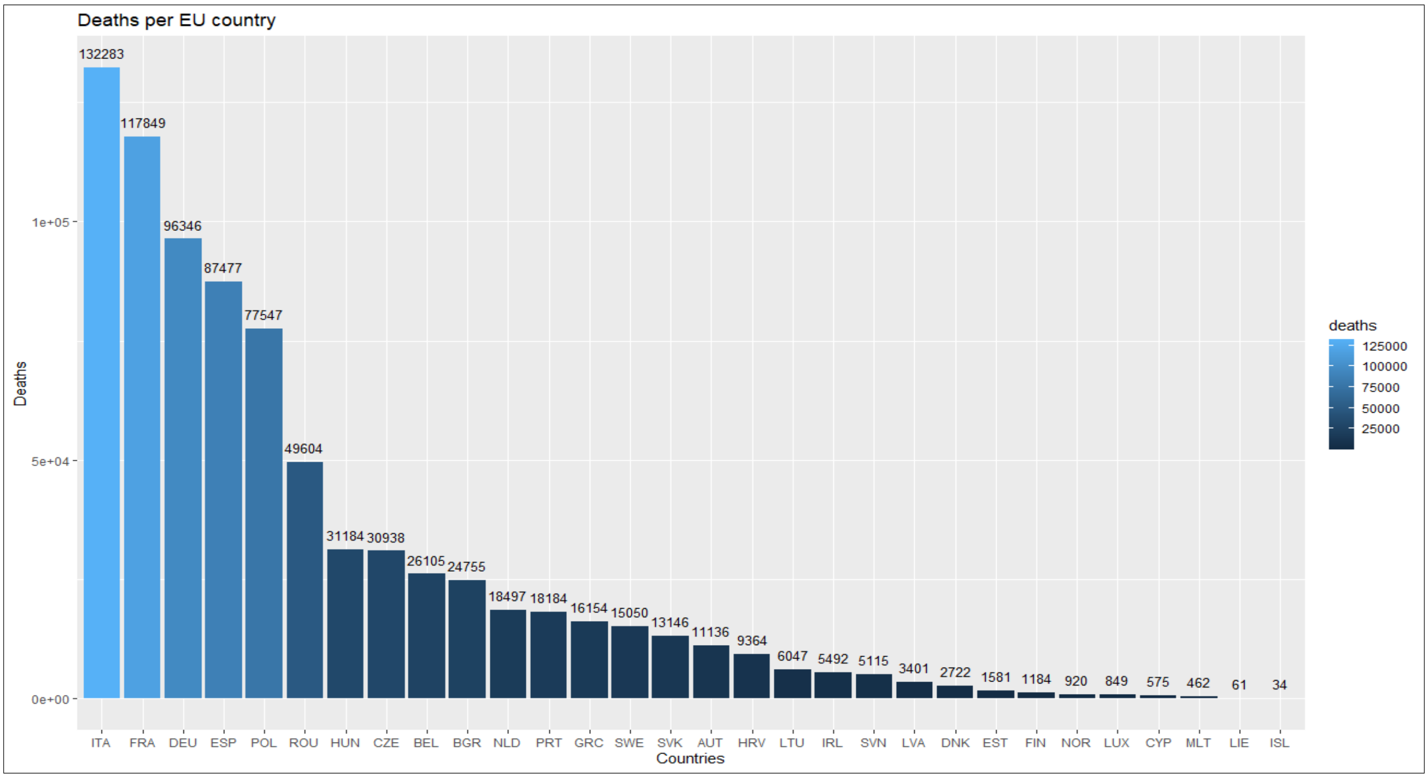
After creating a new dataframe containing the total data for every country, I started to plot some data and to compute some new metrics.

```
> head(cases_tot)
  countriesAndTerritories countryterritoryCode   cases deaths popData2020 deathsPerMille mortalityRate casesPerMille
1                 Austria                  AUT  853310  11136     8901064      1.2510864     1.3050357      95.86607
2                 Belgium                  BEL 1403453  26105    11522440      2.2655792     1.8600552     121.80172
3                Bulgaria                  BGR  623946  24755     6951482      3.5611111     3.9674908      89.75726
4                 Croatia                  HRV  483142   9364     4058165      2.3074468     1.9381465     119.05430
5                  Cyprus                  CYP  125413    575      888005      0.6475189     0.4584852     141.23006
6                 Czechia                  CZE 1801154  30938    10693939      2.8930406     1.7176766     168.42756
```

4.1.1 Covid-19 deaths

First, I plotted the total number of deaths for every EU country.

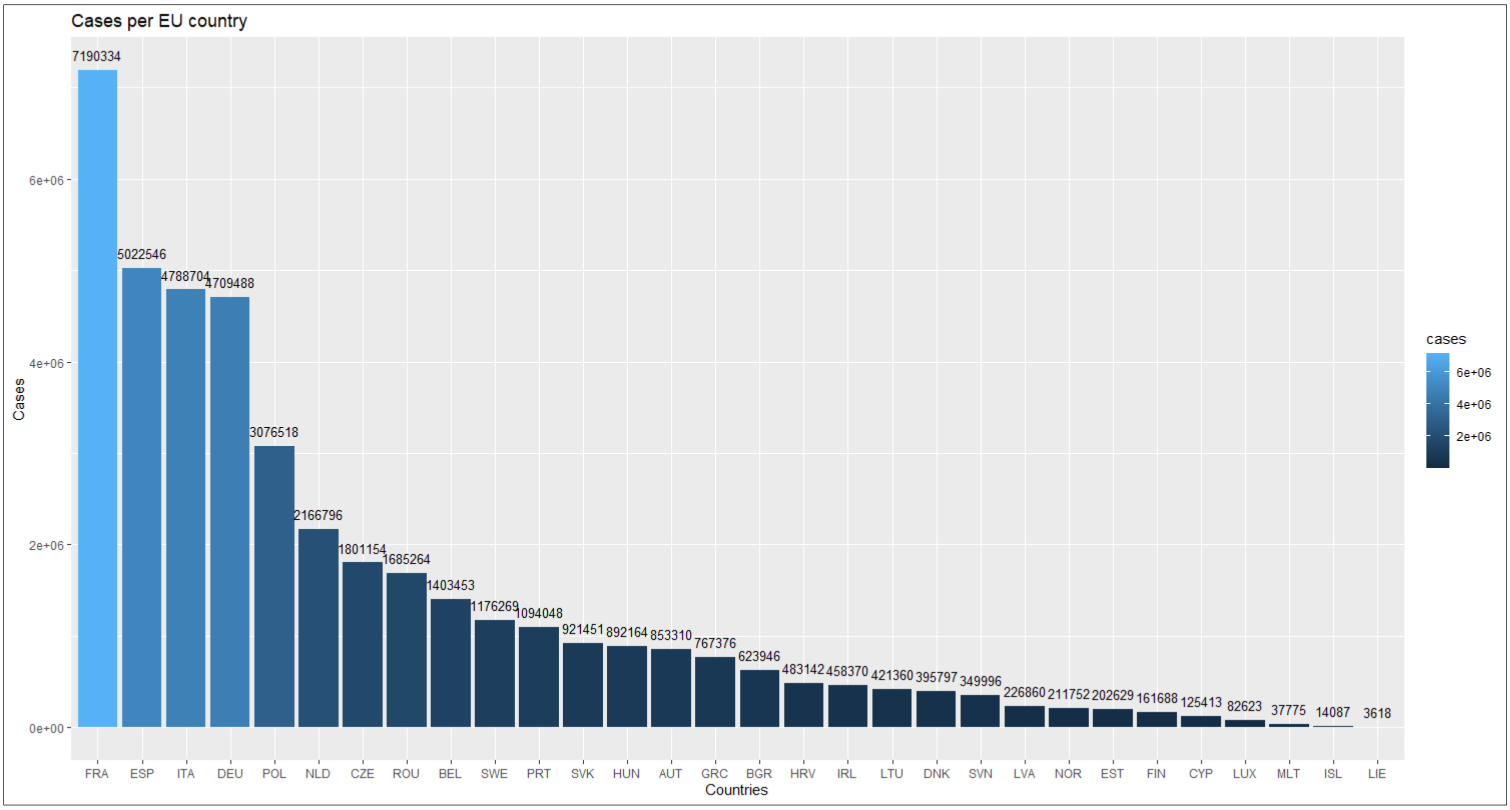
Italy is in first place for Covid-19 related deaths, with 132 283 deaths, followed by France (117 849), Germany (96 346), Spain (87 477) and Poland (77 547).



4.1.2 Covid-19 cases

The second plot I made, represents the total number of cases for every European country.

The country with the most cases as of today has been France, with 7 190 334 people found positive to Covid-19, followed by Spain (5 022 546), Italy (4 788 704), and Germany (4 709 488)



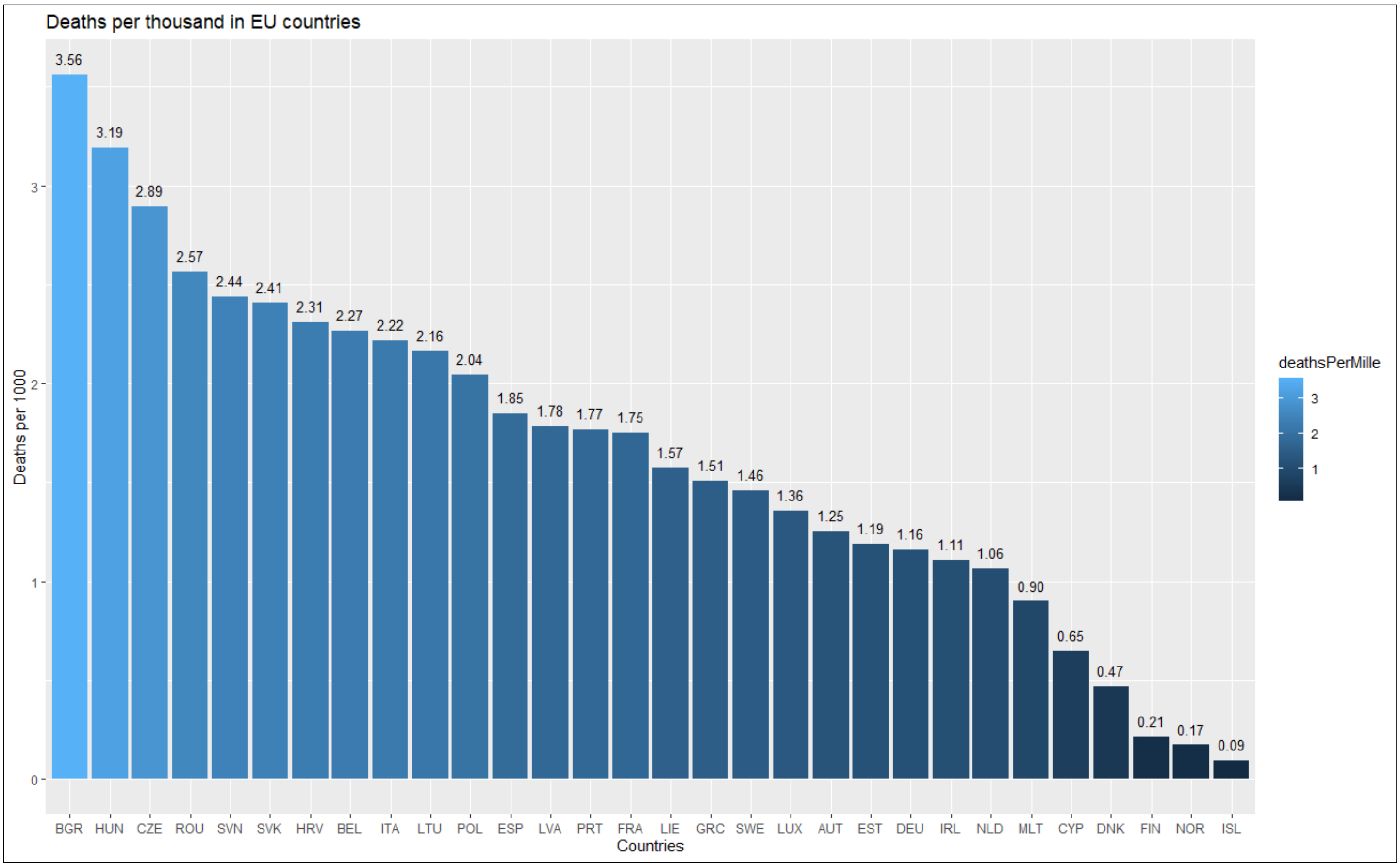
Then, I normalized this data based on population.

4.1.3 Covid-19 deaths per thousand people

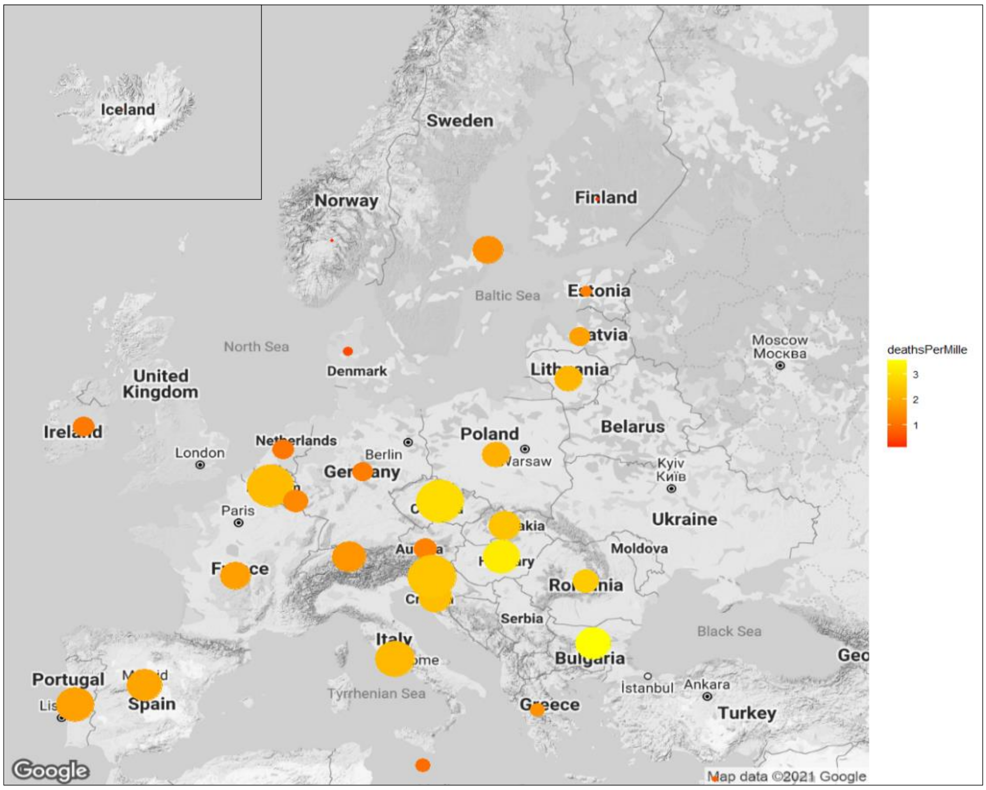
To do this, I used the formula:

$$\text{deathsPerMille} = \frac{\text{deaths}}{\left(\frac{\text{popData2020}}{1000}\right)}$$

The resulting plot shows that the record of deaths per thousand inhabitants is now in Bulgaria, with 3.56 deaths per 1000 inhabitants, followed by Hungary (3.19), the Czech Republic (2.89) and Romania (2.59). Italy is in 9th place with 2.22 deaths per thousand inhabitants.



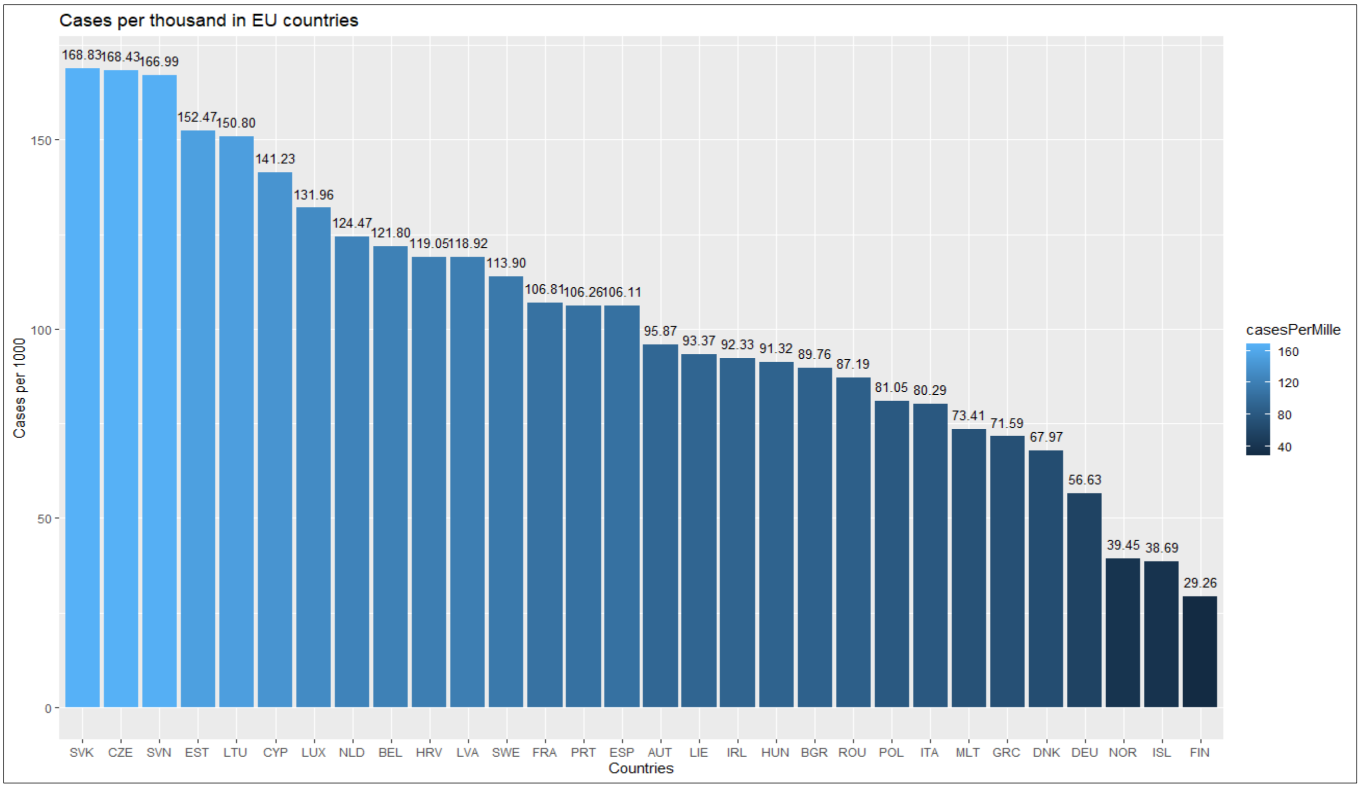
I also plotted the EU map for these metrics, to show which countries struggled the most in the Covid-19 pandemic. Here is the deaths per thousand inhabitants plot:

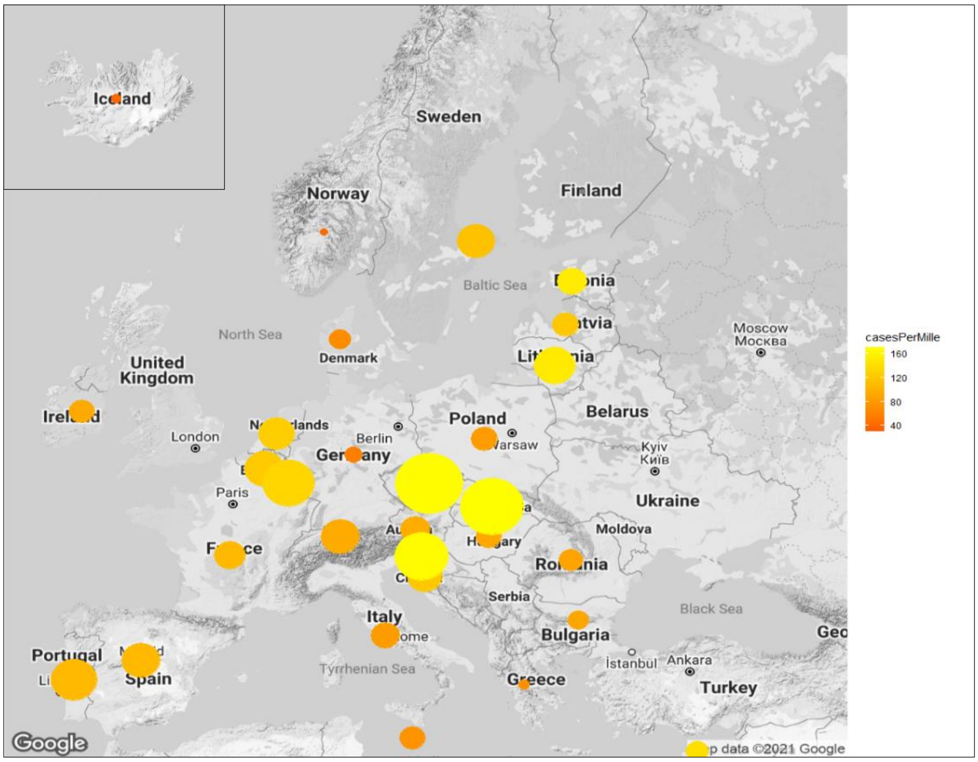


4.1.4 Covid-19 cases per thousand people

Then, in the same way, I computed the number of cases per thousand inhabitants.

The countries with the most cases per thousand people are Slovakia (168.83), Czech Republic (168.43) and Slovenia (166.99)





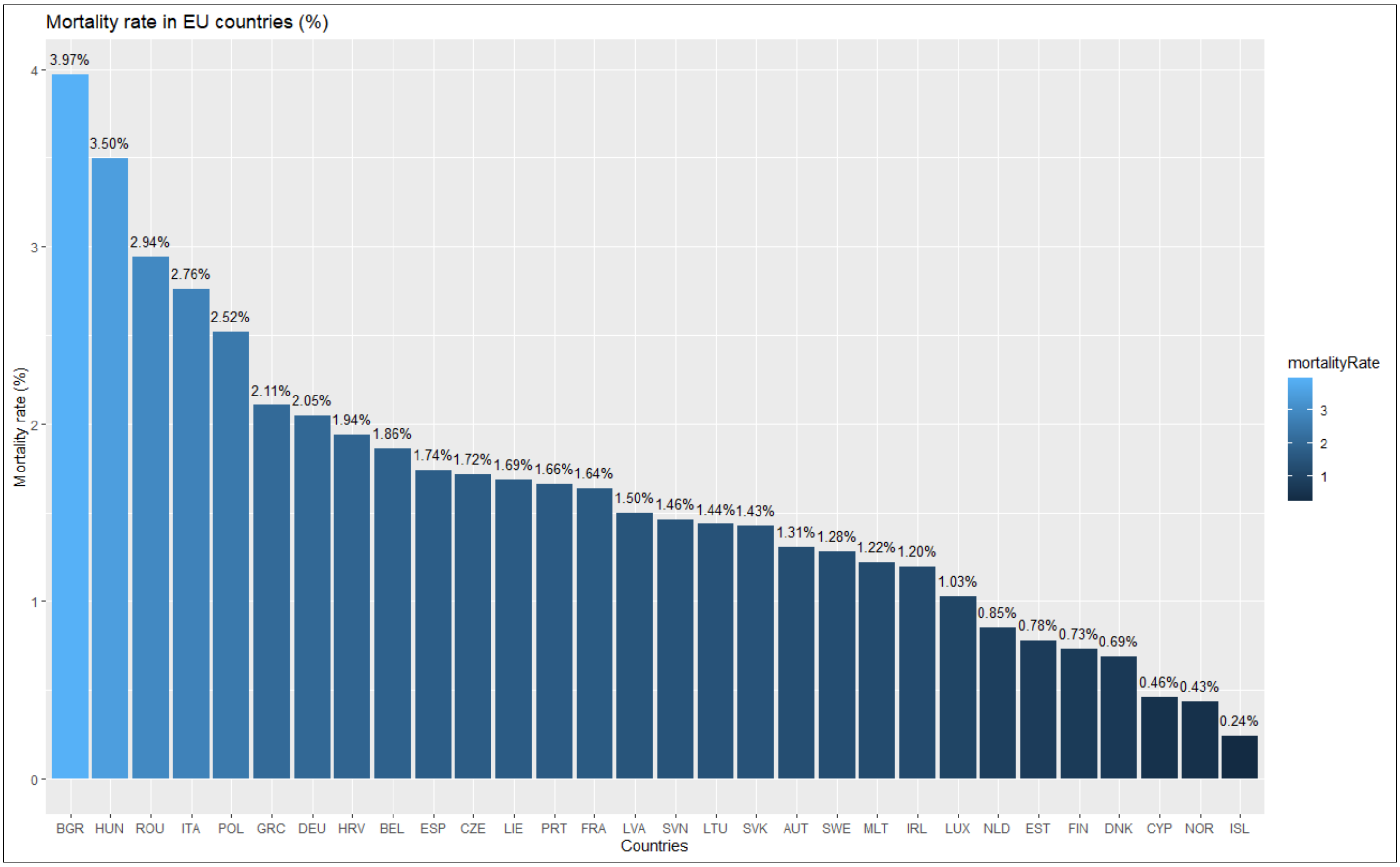
From both maps, it appears that the central Europe (Czech Republic, Slovenia, Hungary, Slovakia) struggled the most with the control of the pandemic.

4.1.5 Mortality rate of Covid-19

I also computed the mortality rate, by doing:

$$\frac{\text{deaths}}{\text{cases}} * 100$$

The plot shows that for the mortality rate, Bulgaria ranks first, with 3.97% of Covid-19 cases which then resulted in death, followed by Hungary (3.50%), Romania (2.94%) and Italy (2.76%).



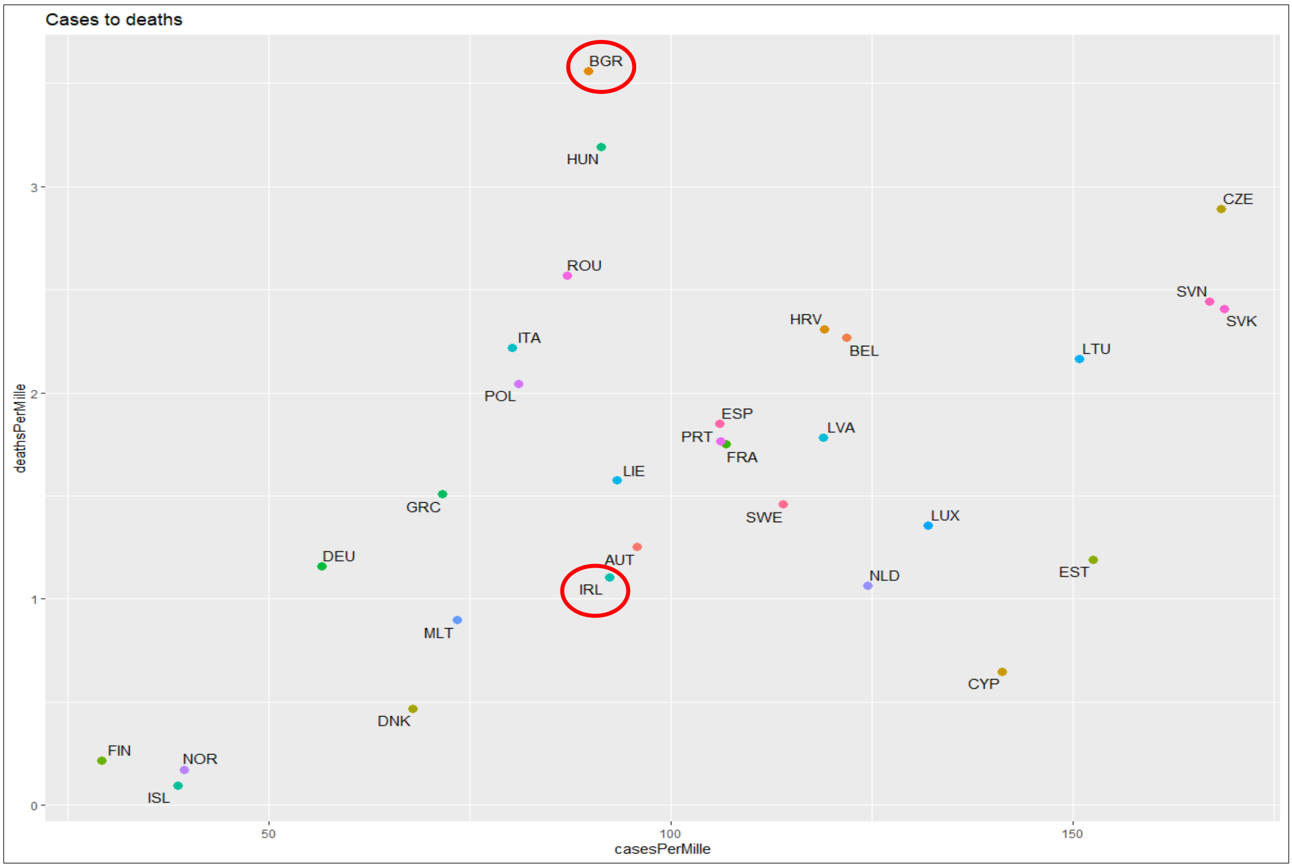
4.1.6 Cases per thousand to deaths per thousand (each EU country response to Covid-19 cases)

Finally, I used these newly computed metrics, to show how well every EU country responded to the rising Covid-19 cases, by plotting a scatter plot.

This plot, clearly shows how different countries with the same cases per thousand inhabitants, got way different deaths per thousand inhabitants (same thing for same deaths and different cases). This difference is mainly attributable to the response of the country's healthcare system, but there may be other factors influencing the outcome.

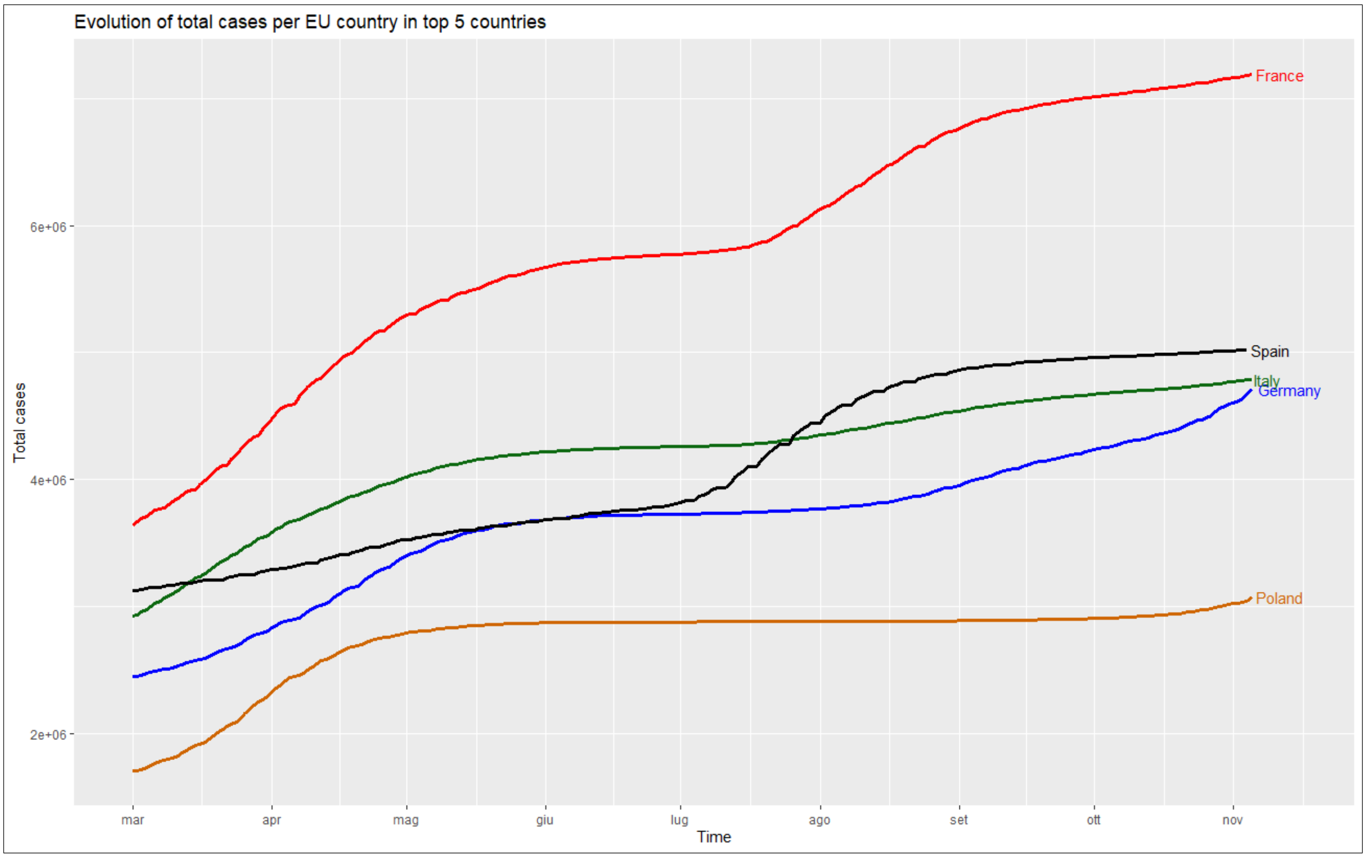
The most striking example of this difference is the Bulgaria-Ireland distance in terms of deaths per thousand inhabitants.

	BULGARIA	**IRELAND**
Cases per thousand	89.76	92.33
Deaths per thousand	**3.56**	1.11
Population	6 951 482	4 964 440
Cases	623 946	458 370
Deaths	**24 755**	5 492



4.1.7 Evolution in time of Covid-19 cases

I tried plotting the evolution of total cases and new cases for each EU country in the dataset (01/03/2021 – 05/11/2021), but the result was a mess due to the overlapping lines, so I kept only the top 5 countries in terms of numbers: France, Spain, Italy, Germany, and Poland.



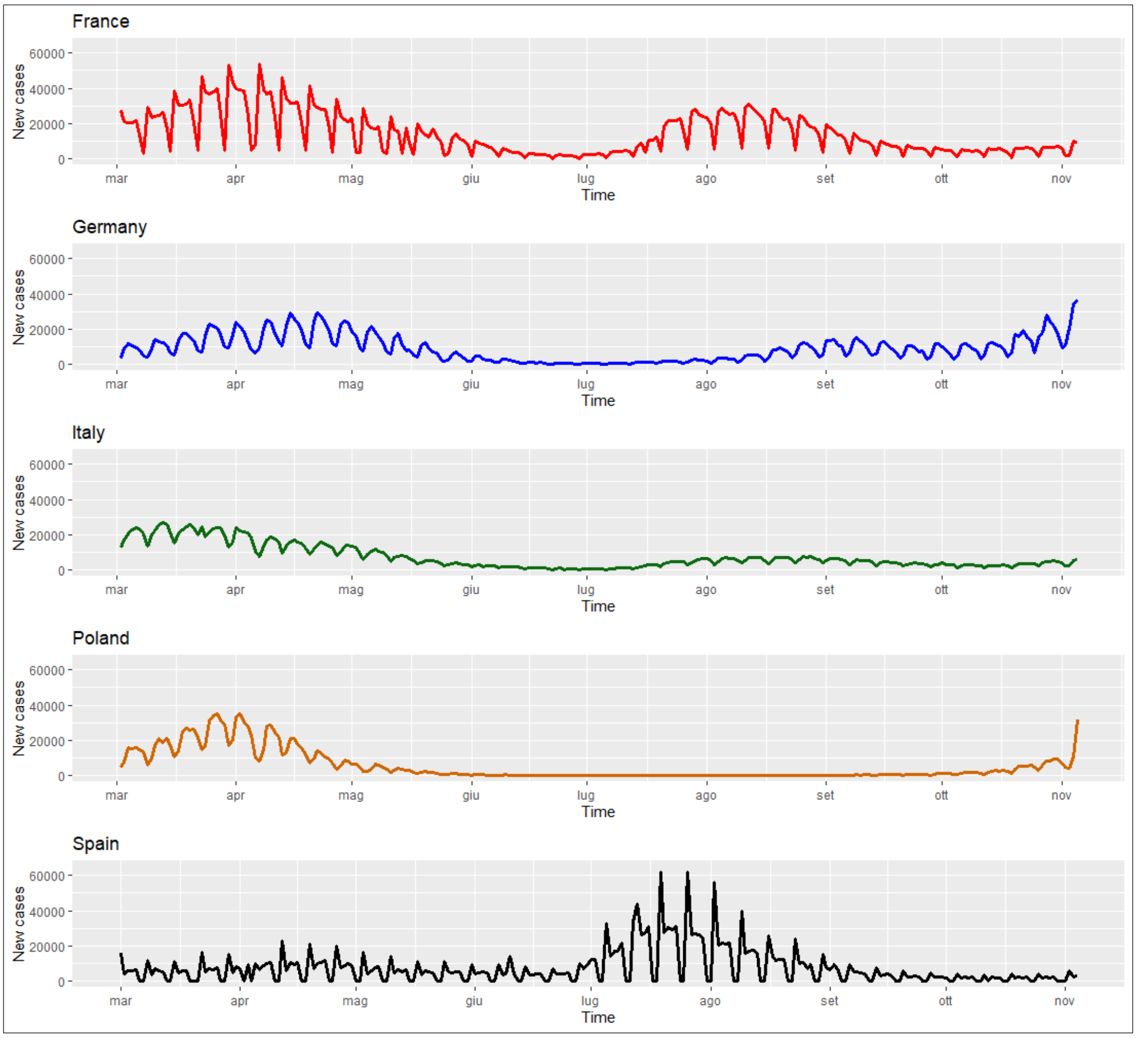
The evolution of total cases is similar for every country. In particular, the data evolution of Spain has a strong surge in July 2021, due, according to the Spanish government, to the Delta variant outbreak, and to the young population not yet vaccinated, which therefore did not cause a significant increase in serious cases and hospitalizations.

The situation is better for Italy and Poland, which during the summer and the arrival of autumn were able to keep the infections under control.

Instead, there is a major surge in cases in Germany with the arrival of the colder months.

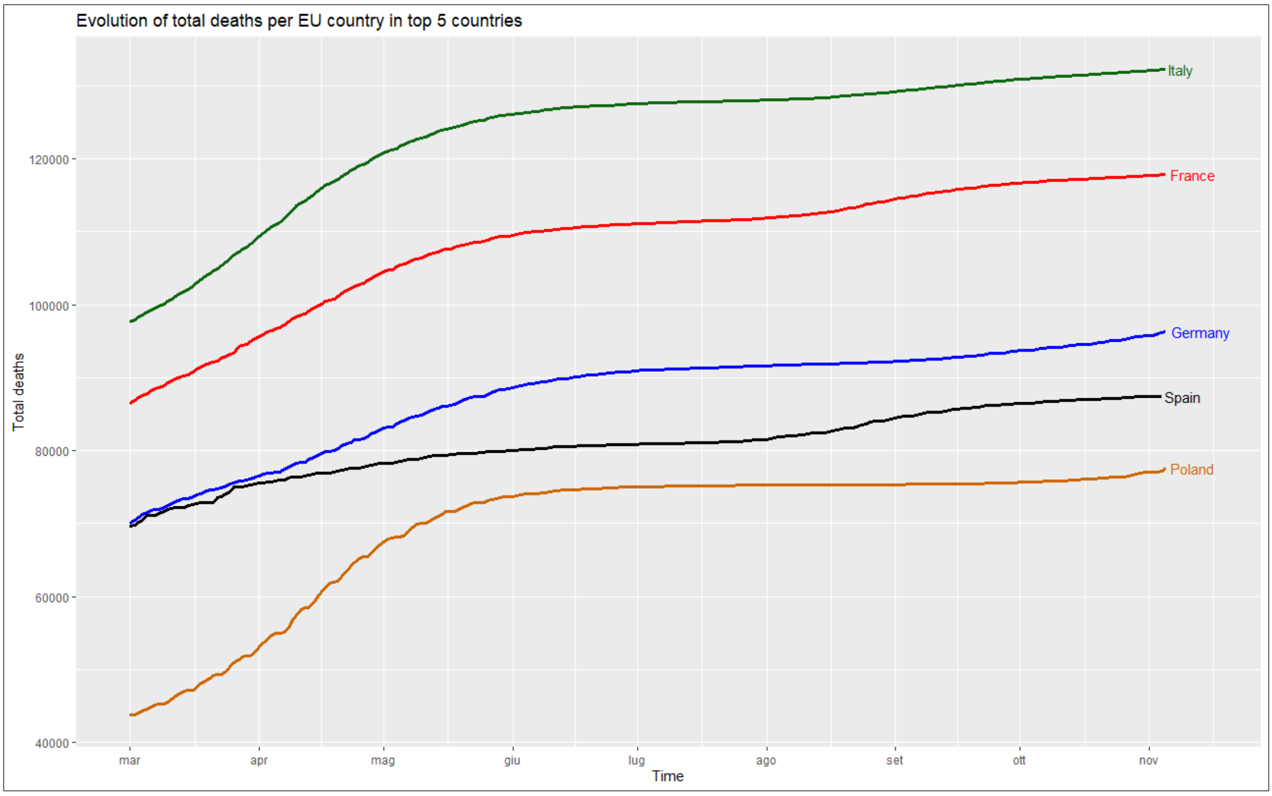
I also plotted the new daily cases. To do this, I had to eliminate the 1st of March data, because it contains the total numbers from previous months, and would have made the plot unreadable.

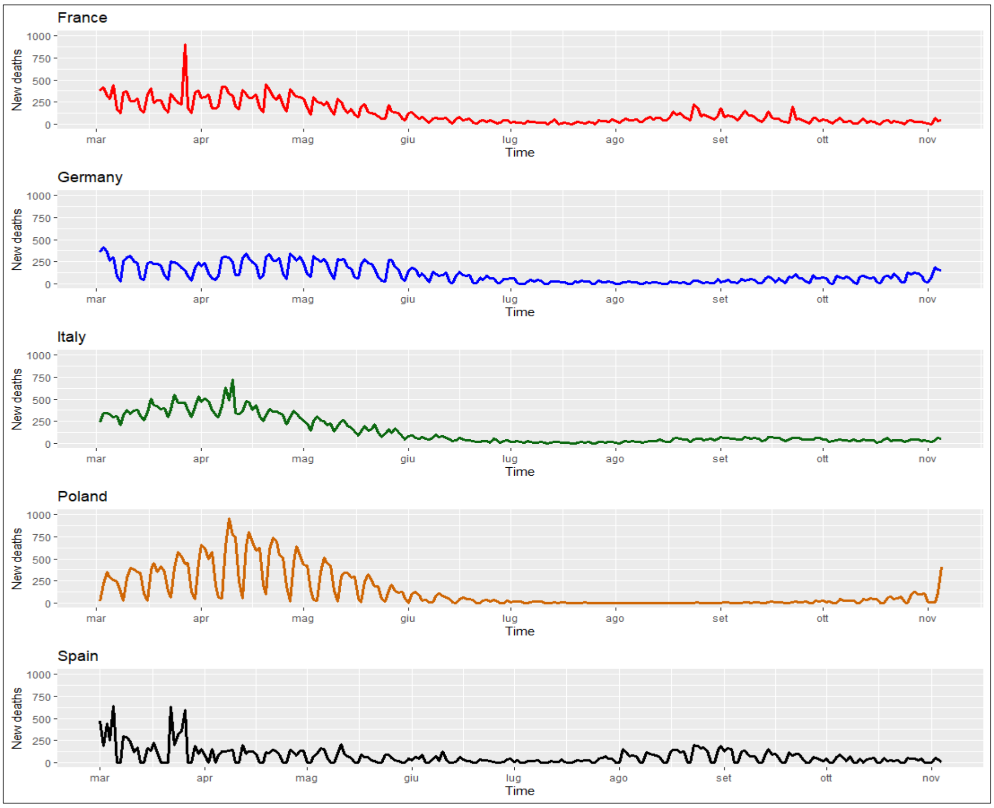
This plot confirms what is represented in the first one. The strong wave of July for Spain, and the surge of cases for Germany.



4.1.8 Evolution in time of Covid-19 deaths

I plotted the same two graphs for Covid-19 deaths. The result shows that the deaths trend is no longer exactly following that of cases, a symbol that the infected are able to recover and heal more easily.

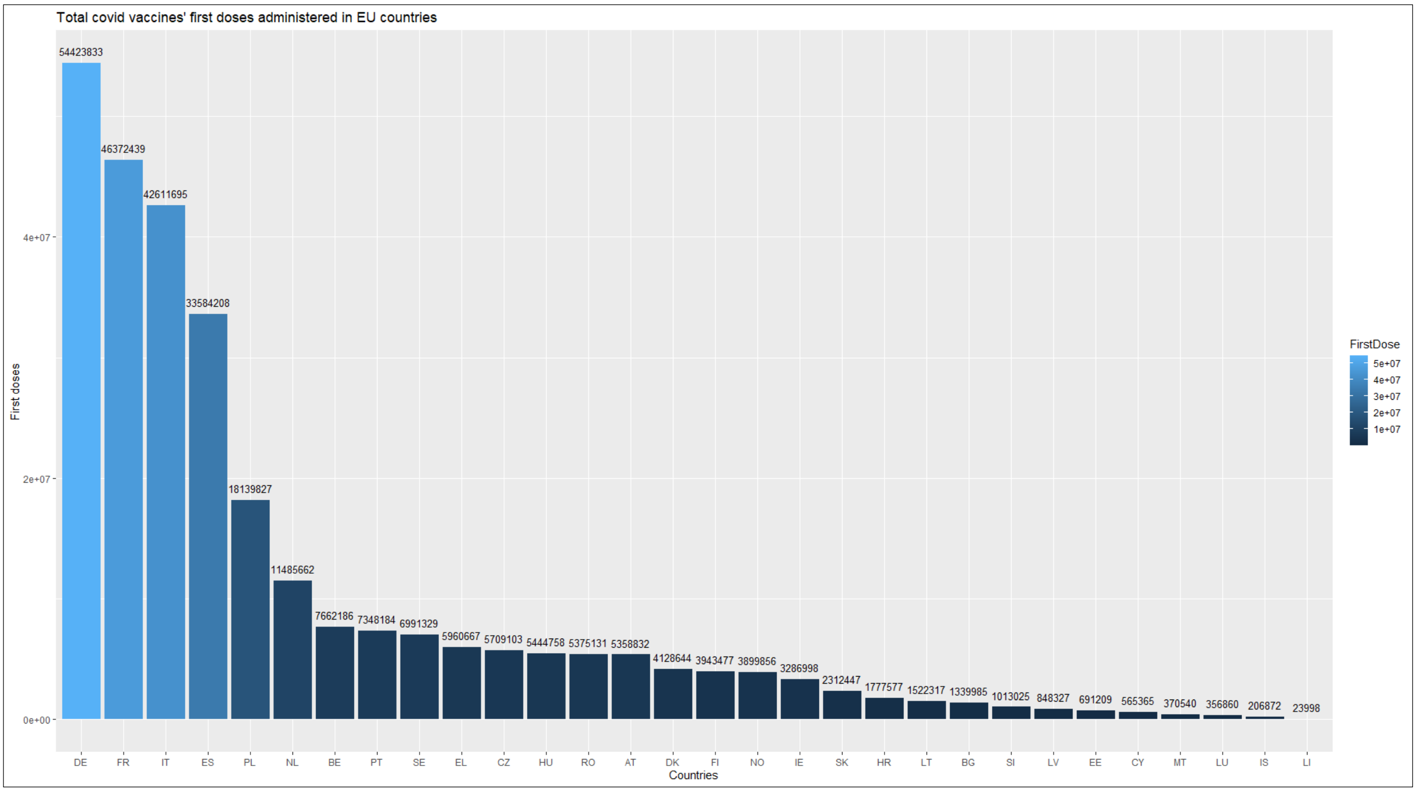


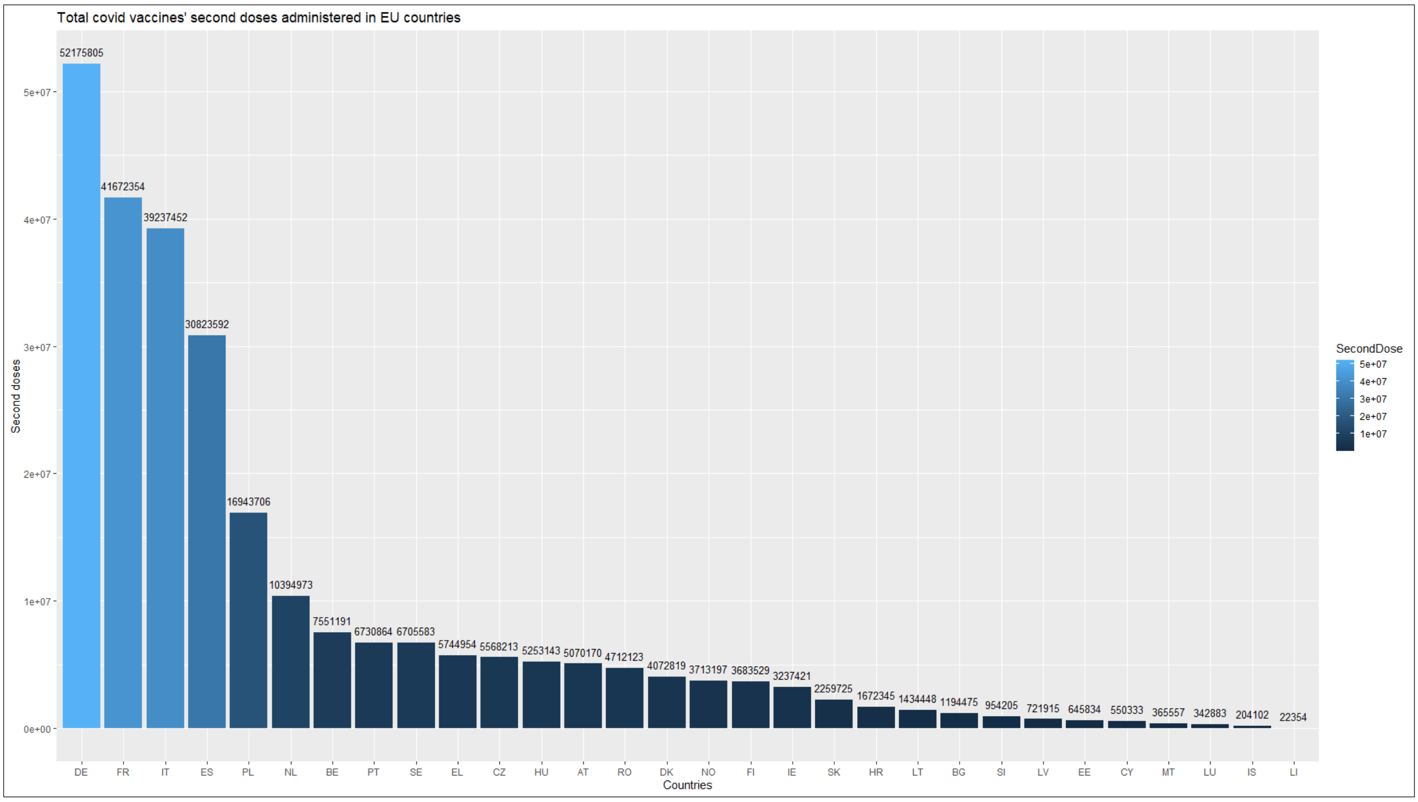


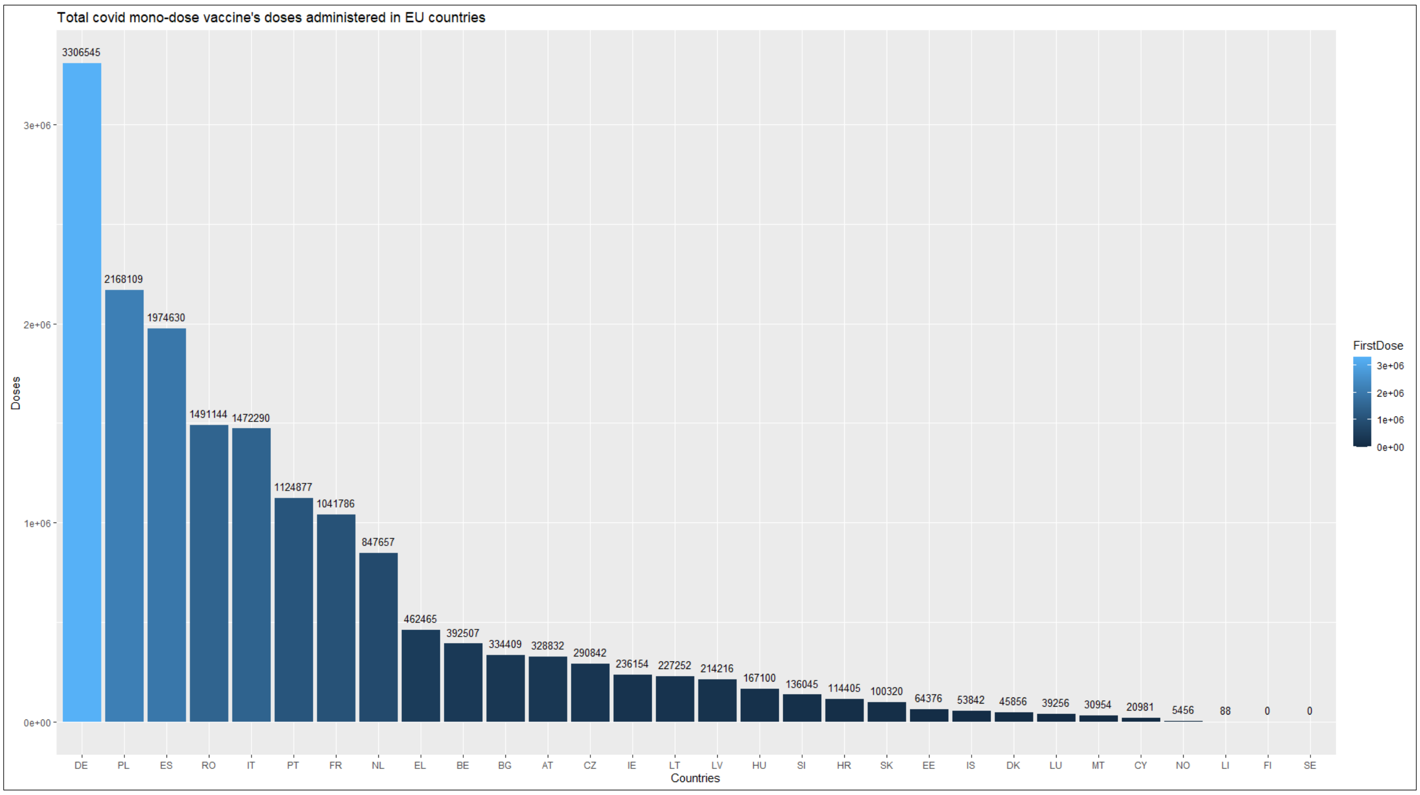
4.2 VACCINES DATASET

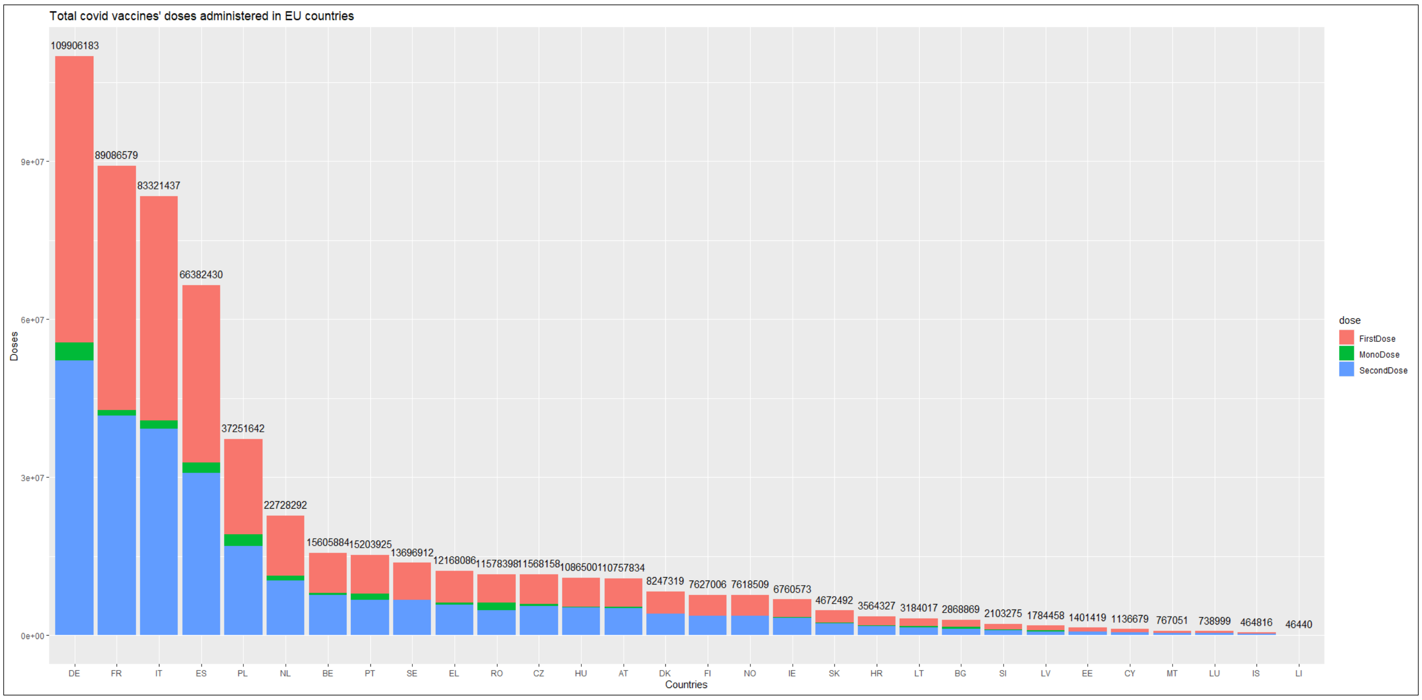
4.2.1 Total number of doses

First, I plotted the total number of every dose for each EU country:





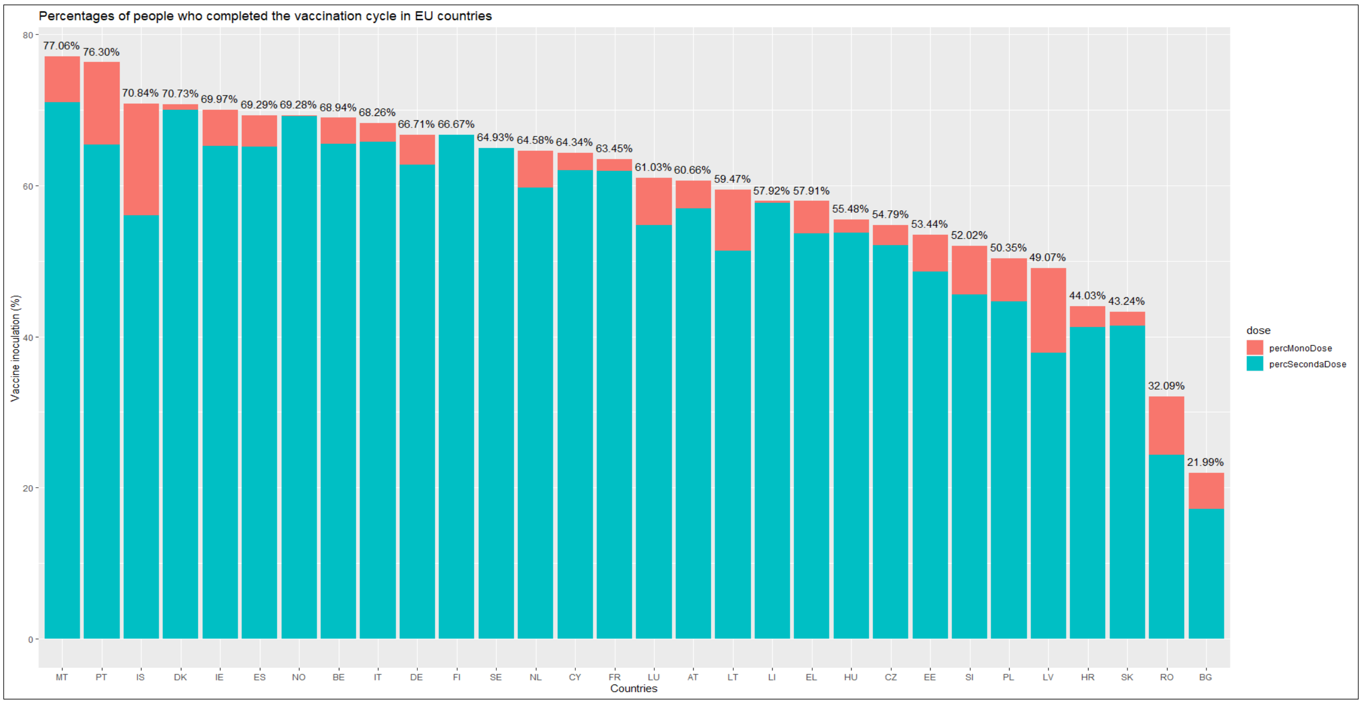


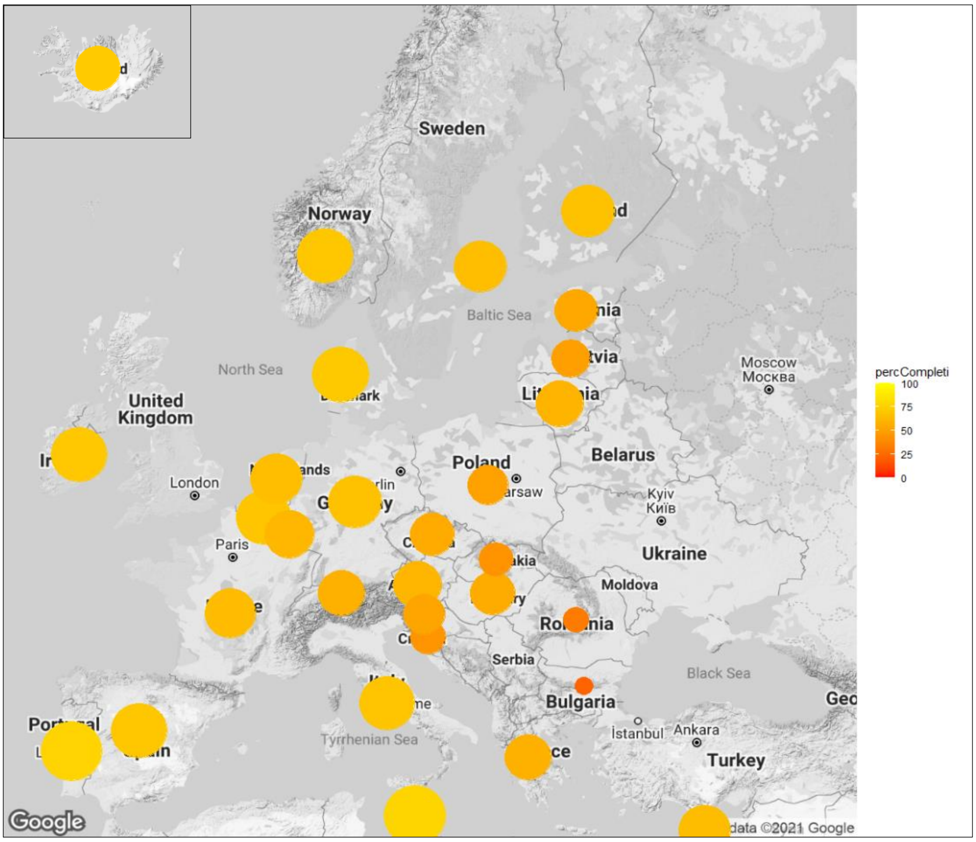


The country with the most administered vaccine doses is Germany, with a total of 109 906 183 vaccines. It is followed by France (89 086 579 doses), Italy (83 321 437 doses), and Spain (66 382 430 doses).

4.2.2 Percentage of people who completed the vaccination cycle

Then, I have calculated the percentage of people over the total population who has completed the vaccination cycle. I had to distinguish between the single dose alternative and the second dose of the double dose alternative. Here are the total percentages plotted in a bar chart and in an EU map:



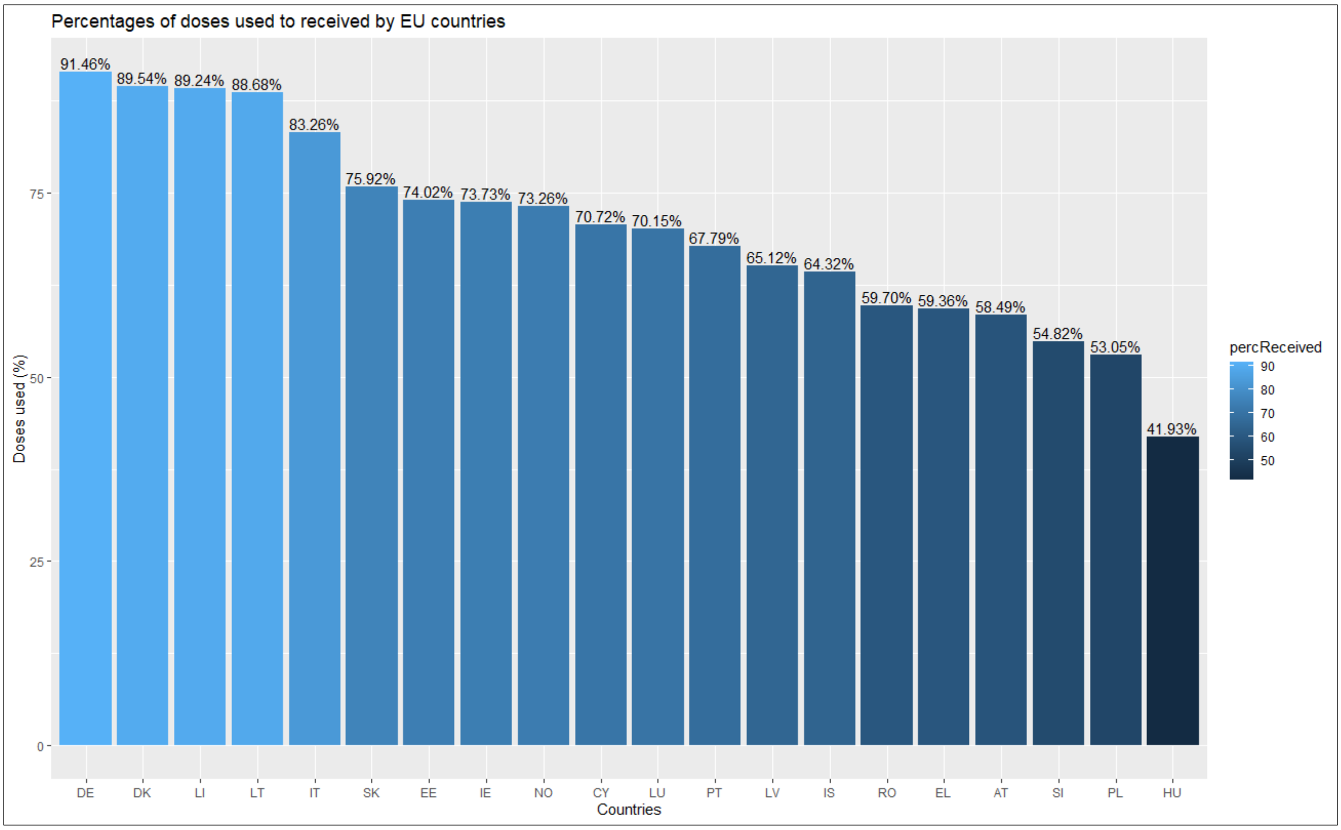


The country with the most fully vaccinated people over the total population is Malta, with the 77.00% of the total population immunized, followed by Portugal (76.30%), Iceland (70.84%), and Denmark (70.73%). Italy is in 9th place with the 68.20% of the population fully vaccinated.

4.2.3 Percentage of doses used compared to those received

Next thing, I computed the percentage of doses used compared to those received, to understand the effectiveness of the vaccination campaign in each EU country.

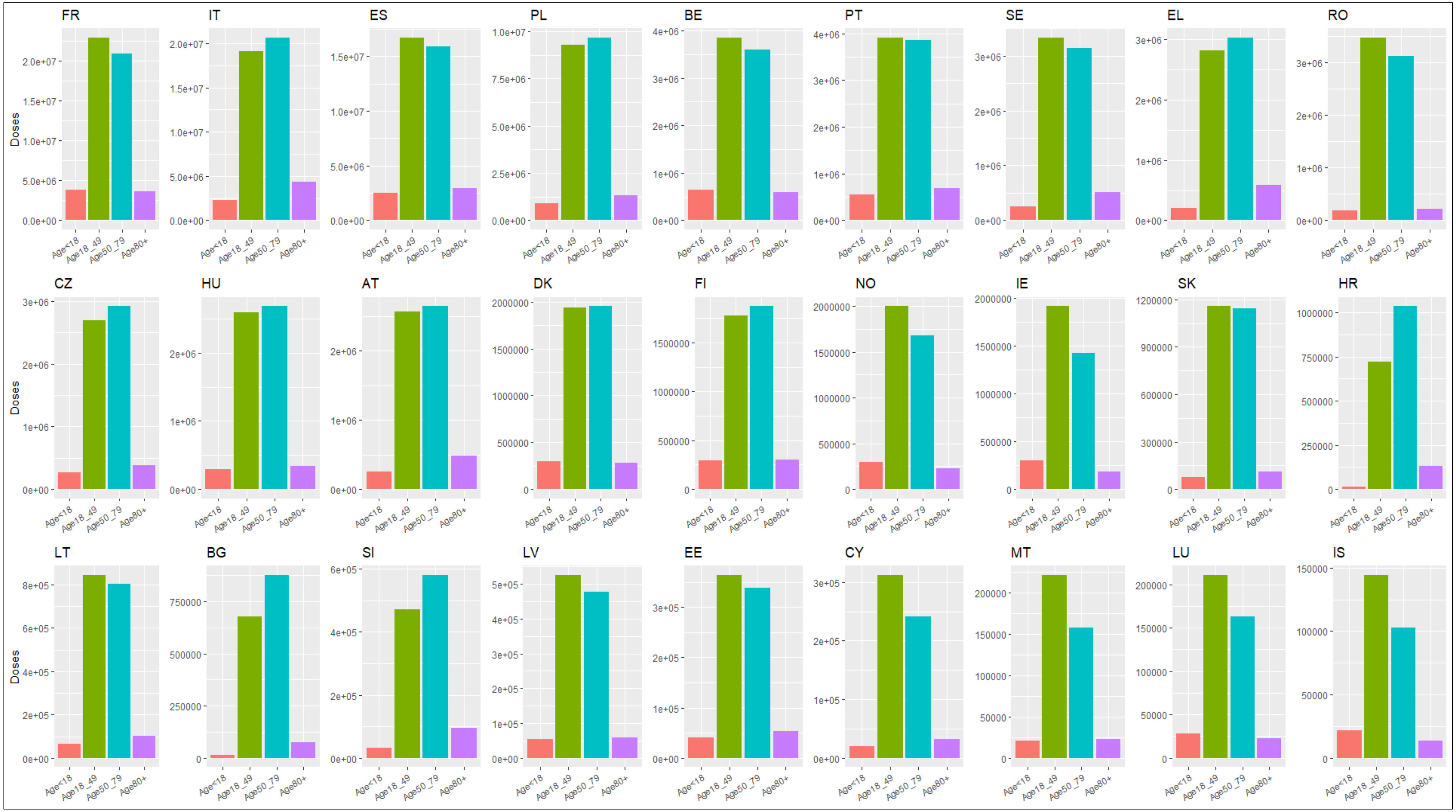
Unfortunately, the data isn't available for every country, so I only plotted the ones available.



4.2.4 Vaccine doses by age group

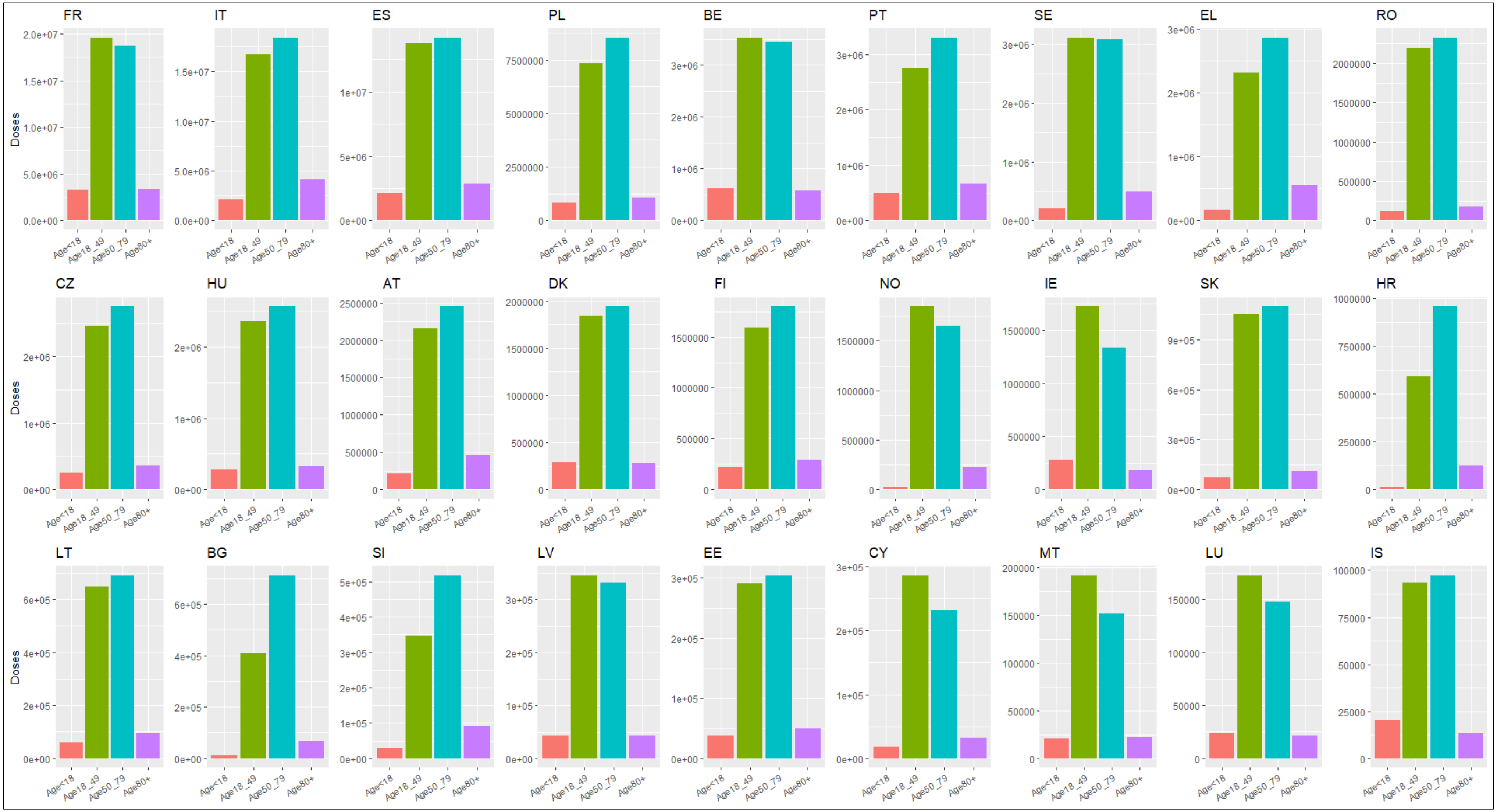
I used the age group classification I created in the data cleaning process, to plot how many doses were administered for every group in each country. I wanted to plot the percentage for every age group, but the group population data was missing in a lot of countries.

FIRST DOSE



Note: the data of the single dose vaccine was included in the first dose plot

SECOND DOSE


FR
IT
ES
PL
BE
PT
SE
EL
RO
CZ
HU
AT
DK
FI
NO
IE
SK
HR
LT
BG
SI
LV
EE
CY
MT
LU
IS
Doses
Age<18
Age18_49
Age50_79
Age80+

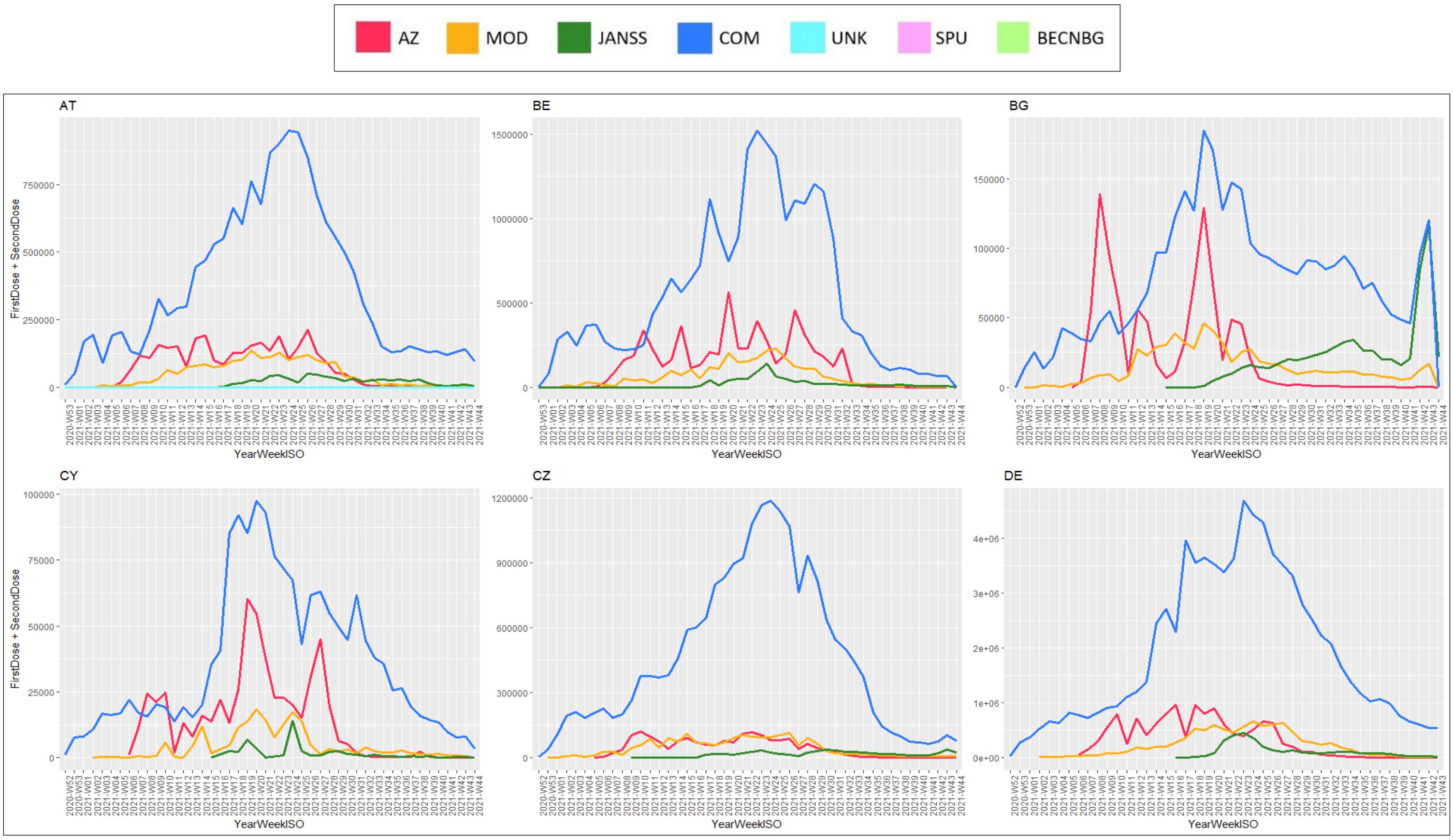
The division by age group, clearly shows that the majority of the vaccine’s doses have been administered to the age 50 to age 79 group. Thought in some cases the Age18_49 column is higher, considering the shorter age gap from 50 to 79, the percentage of vaccinated people in the third age group should be higher.

4.2.5 Evolution of vaccines usage in the vaccination campaign

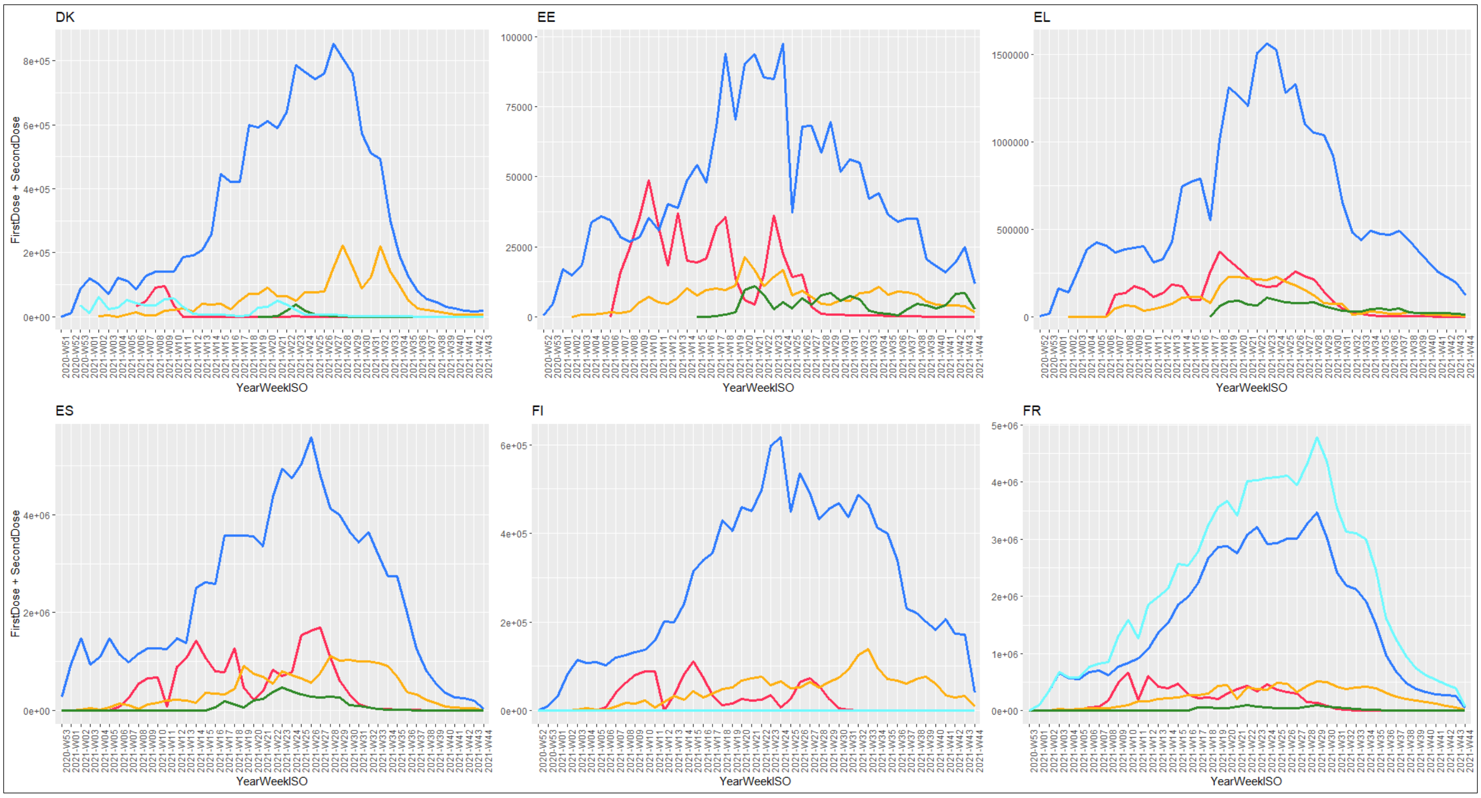
Last thing, I plotted the evolution in time of the usage of the various types of vaccine available for use in Europe.

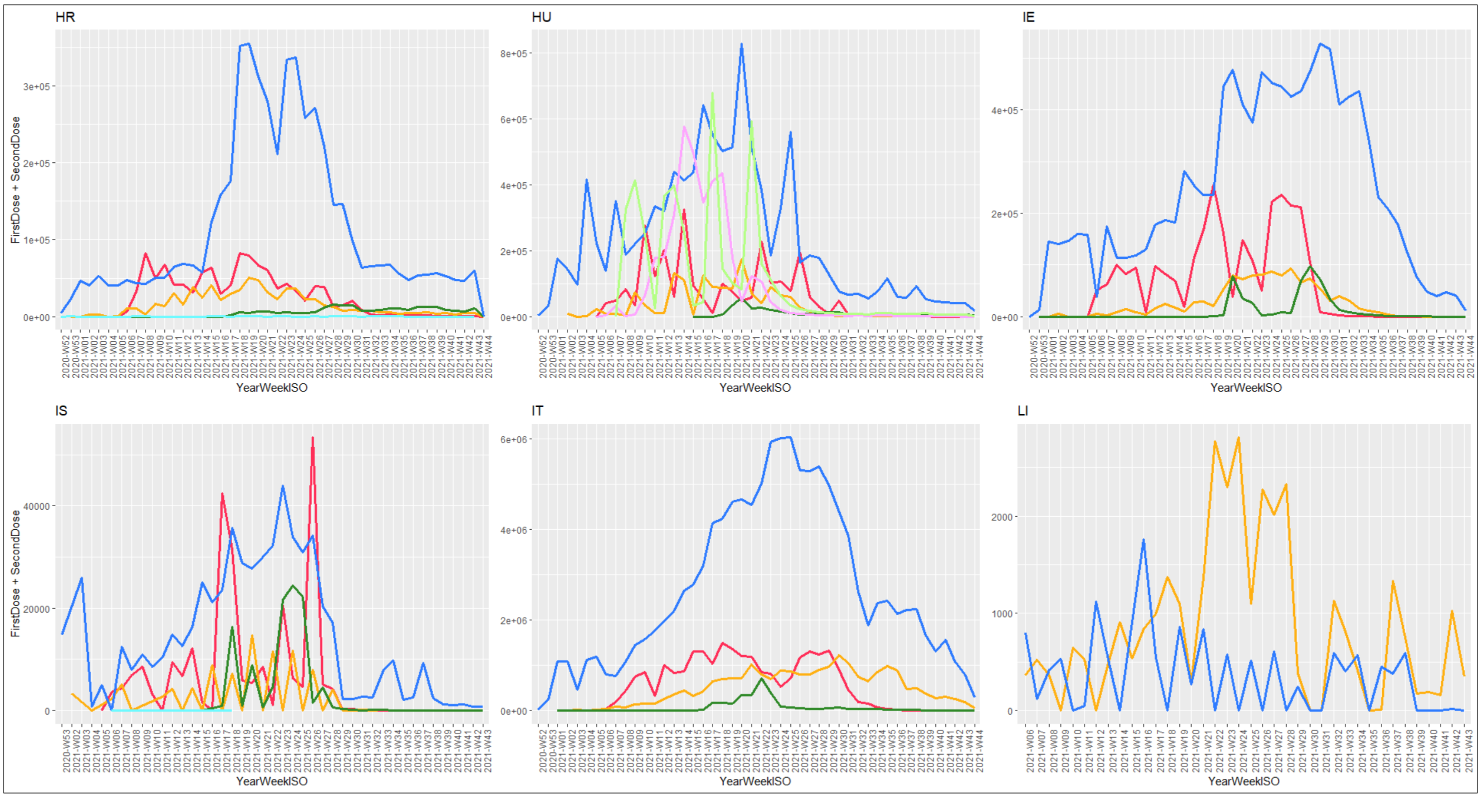
In every EU country, the most used vaccine was Comirnaty (BioNTech, Pfizer), except for Liechtenstein where the Moderna vaccine was more used, and France, where there is a high value of administered vaccine from unknown source (the second one is still Comirnaty).

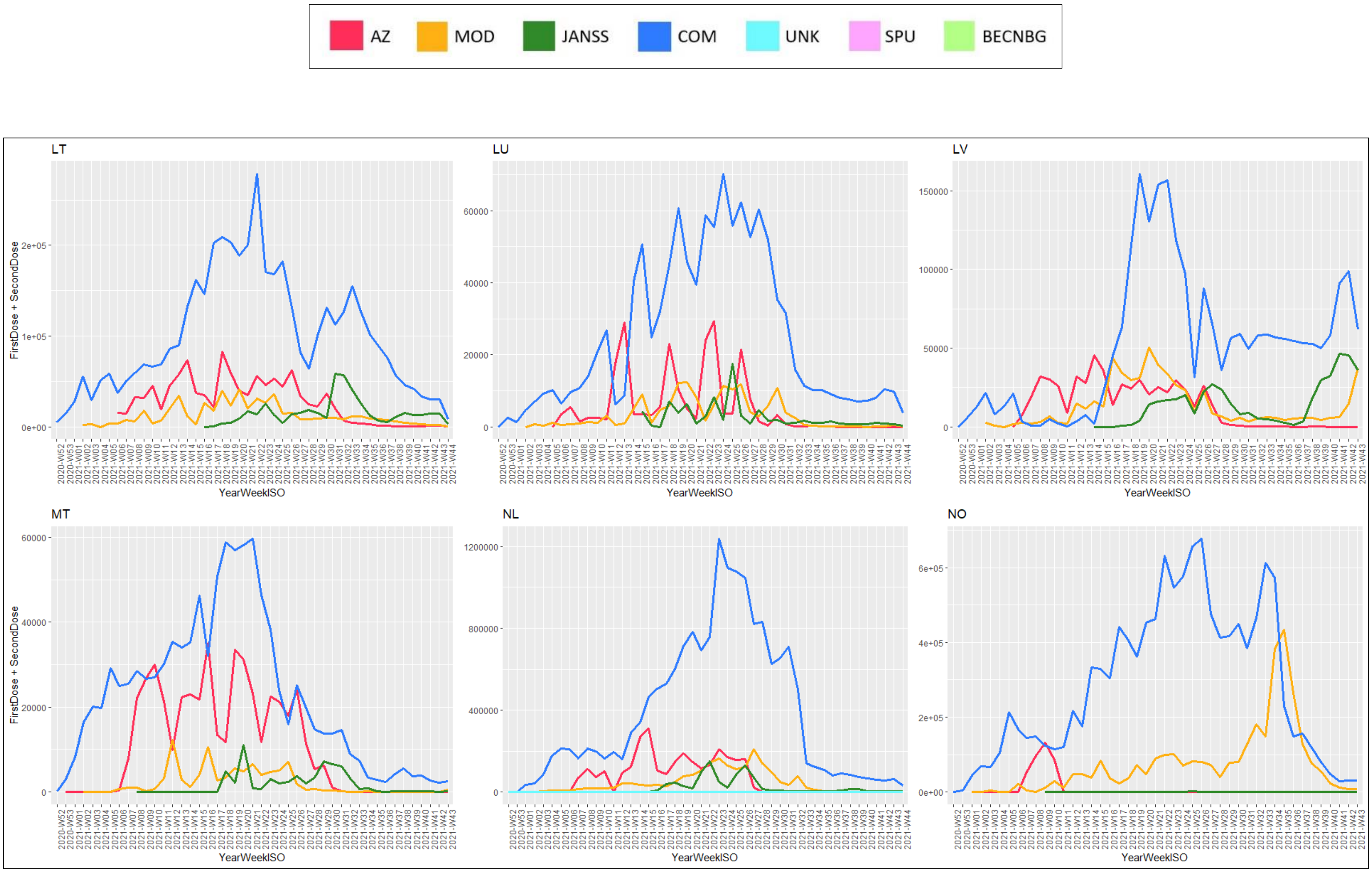
Also, for almost every country, the peak of usage was between the weeks 12/15 and 32/35, so during the spring/summer, because the vaccination campaigns in every country opened to more and more people as time passed, allowing everyone to get the shot by the end of the summer.

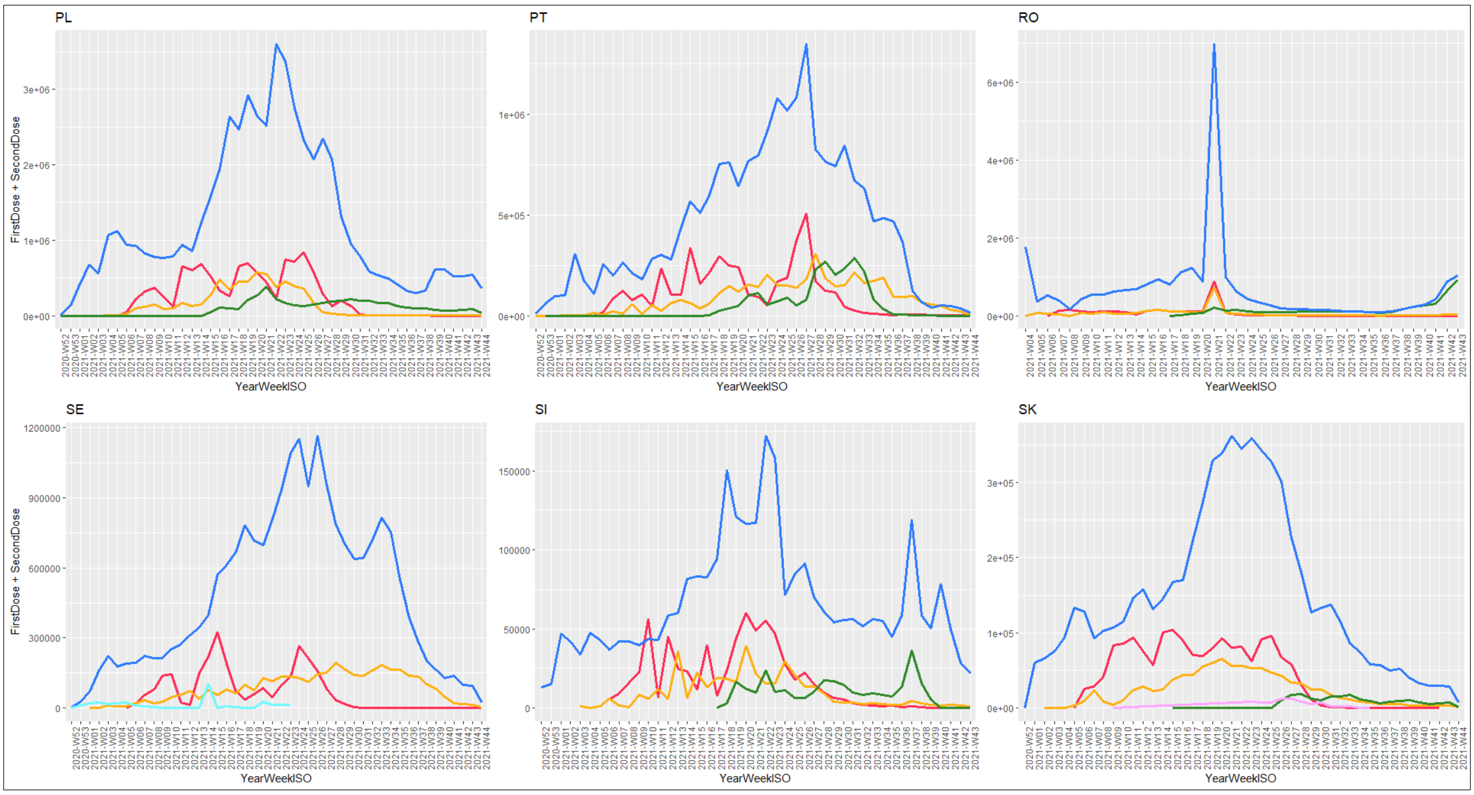



AZ
MOD
JANSS
COM
UNK
SPU
BECNBG


DK
EE
EL
ES
FI
FR
FirstDose + SecondDose
YearWeekISO


HR
HU
IE
IS
IT
LI
FirstDose + SecondDose
YearWeekISO


AZ
MOD
JANSS
COM
UNK
SPU
BECNBG
LT
LU
LV
MT
NL
NO
FirstDose + SecondDose
YearWeekISO


PL
PT
RO
SE
SI
SK
FirstDose + SecondDose
YearWeekISO

5. CASES AND VACCINES COMPARISON

To study the impact of the vaccine campaign on the pandemic evolution, I joined the dataset using the country column as key.

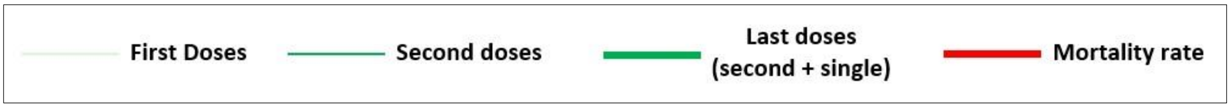
5.1 Evolution of administered vaccines and mortality

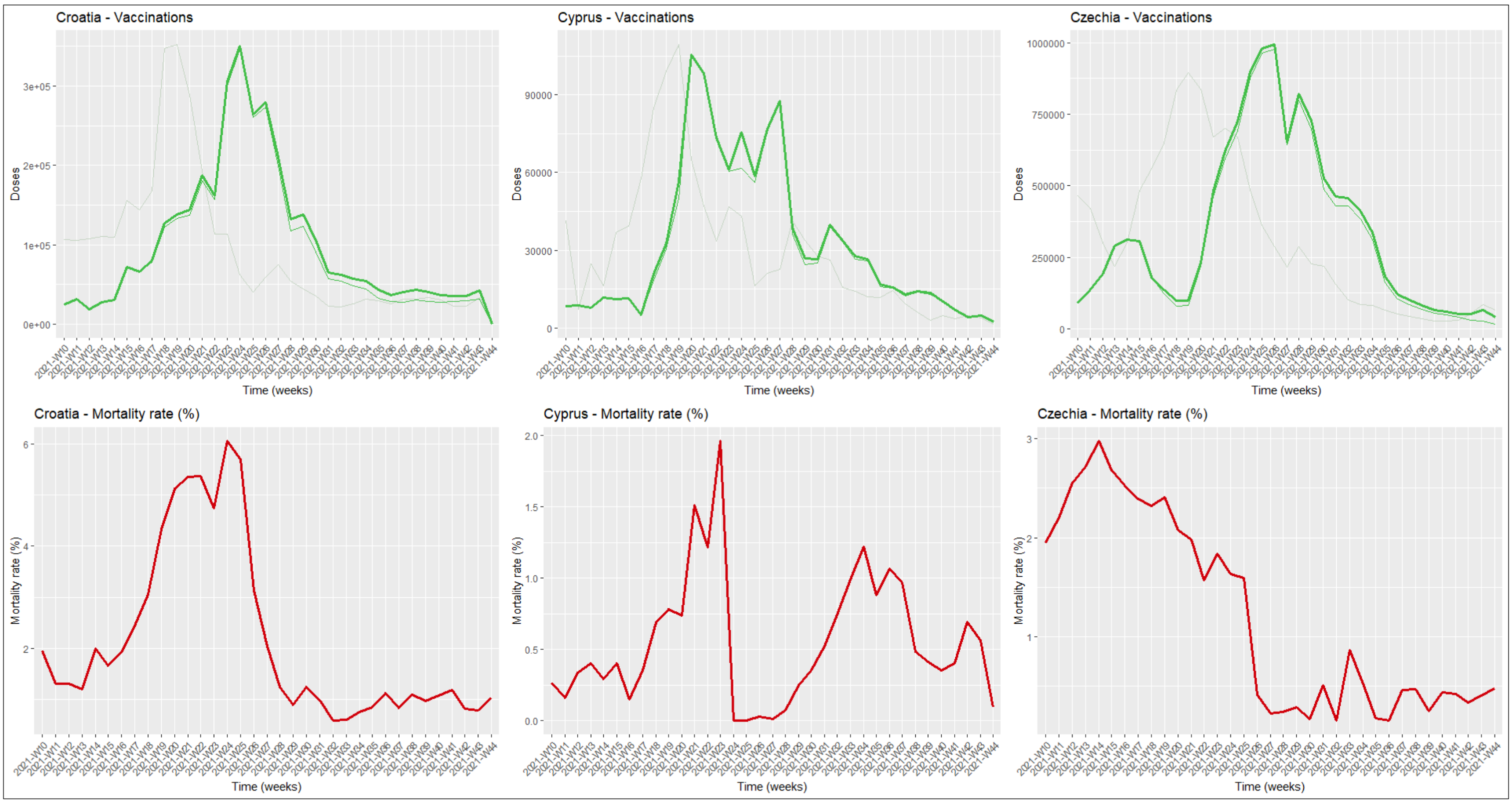
The first comparison I made, was between the evolution of the vaccine campaign, and the evolution of the mortality rate of Covid-19.

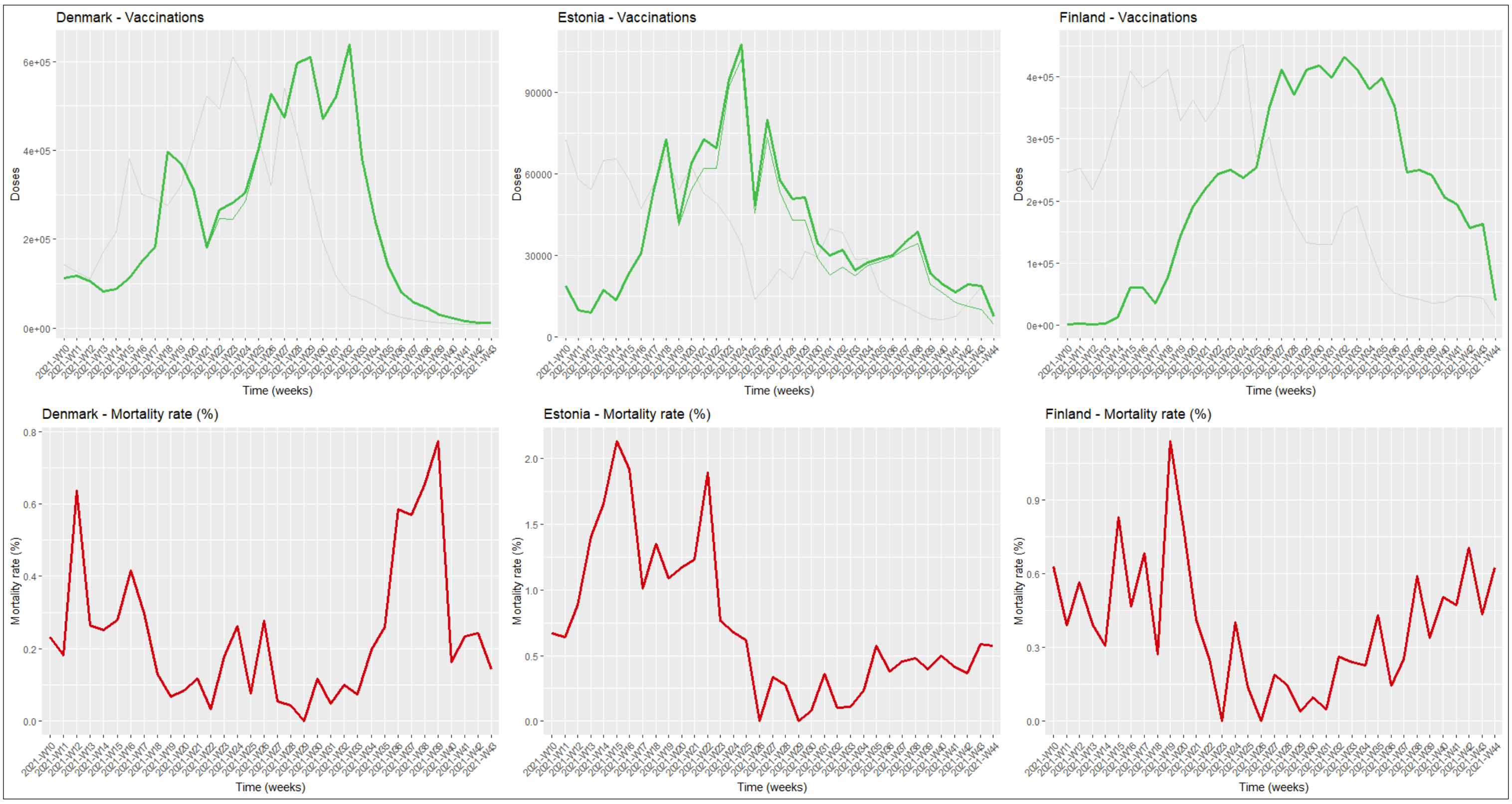
The following plots represent, for each country, a line graph with three different lines:

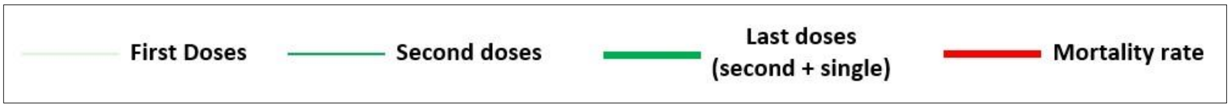
- The thin, dark green line, is the evolution of the administration of the first dose
- The thin, light green line, is the evolution of the administration of the second dose
- The thicker green line is the evolution of the fully immunizing doses administered (second dose + single dose vaccines)

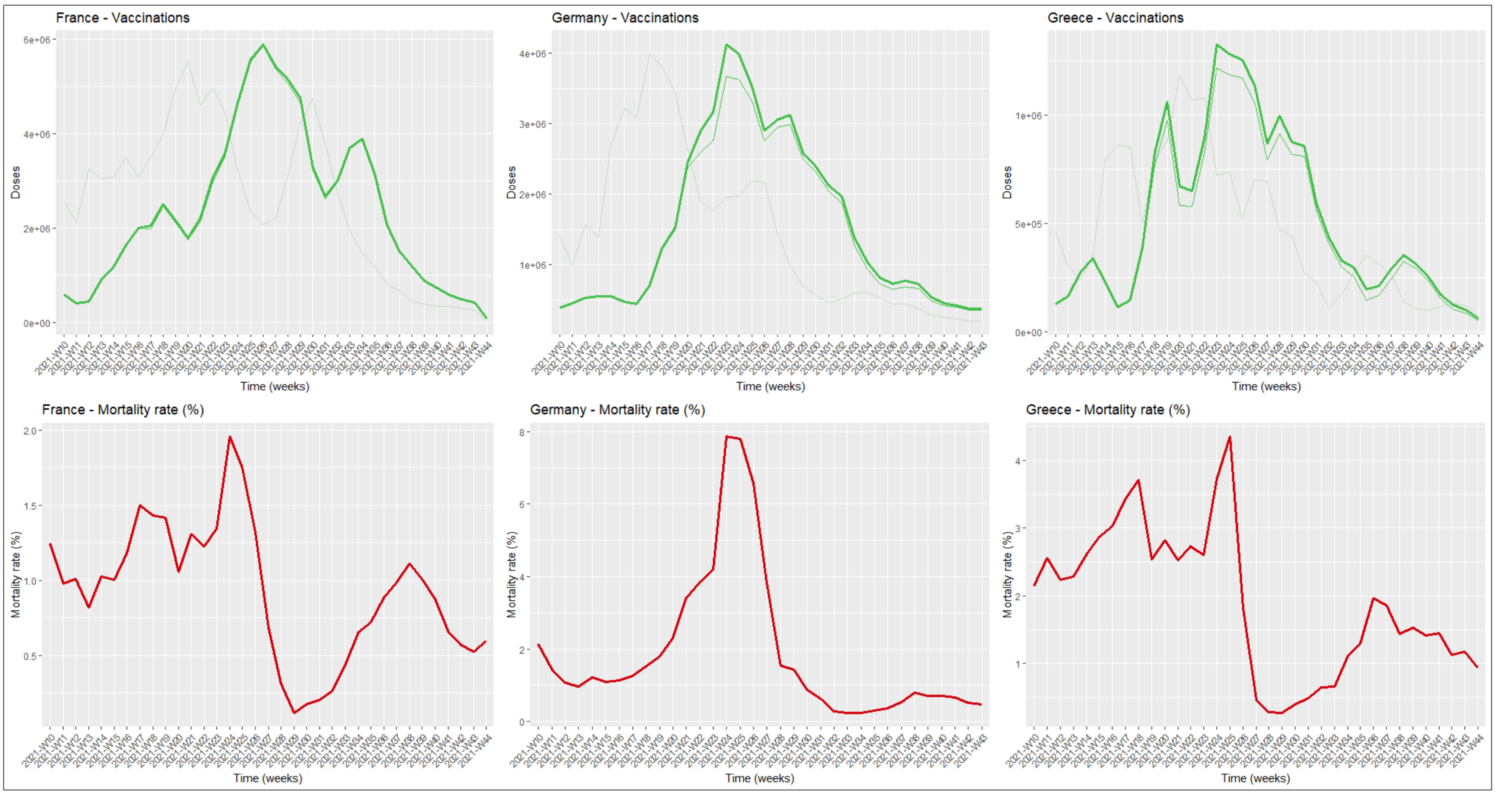
Below this graph, is a line graph showing the evolution of the mortality rate in the same time span.


First Doses
Second doses
Last doses (second + single)
Mortality rate

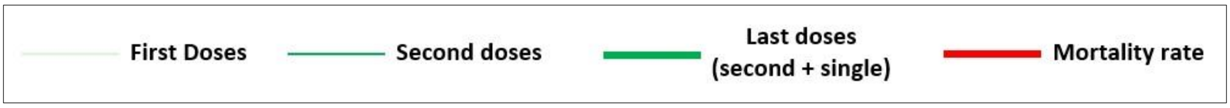

Croatia - Vaccinations
Cyprus - Vaccinations
Czechia - Vaccinations
Croatia - Mortality rate (%)
Cyprus - Mortality rate (%)
Czechia - Mortality rate (%)
Doses
Mortality rate (%)
Time (weeks)

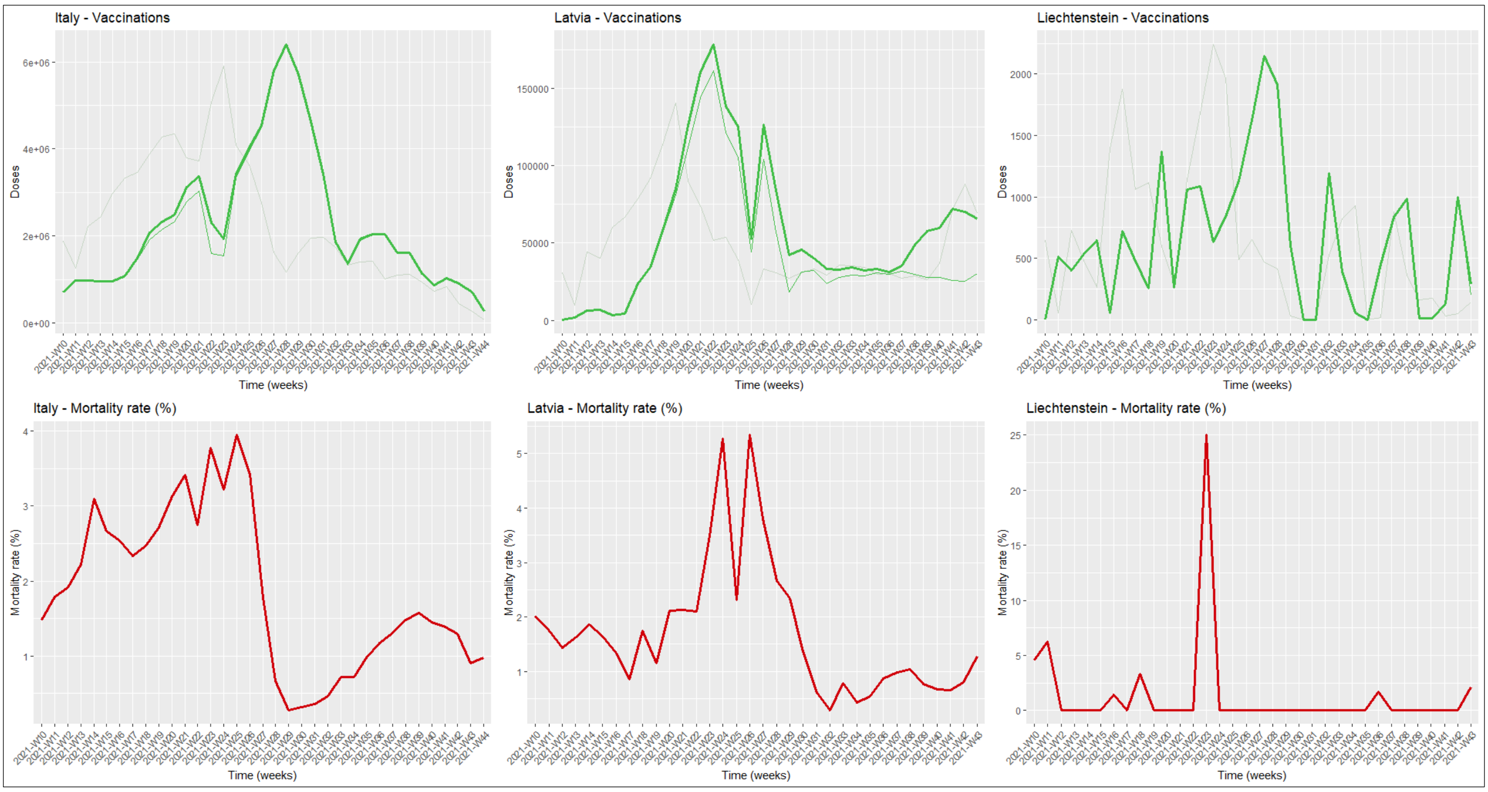

Denmark - Vaccinations
Estonia - Vaccinations
Finland - Vaccinations
Denmark - Mortality rate (%)
Estonia - Mortality rate (%)
Finland - Mortality rate (%)
Doses
Mortality rate (%)
Time (weeks)

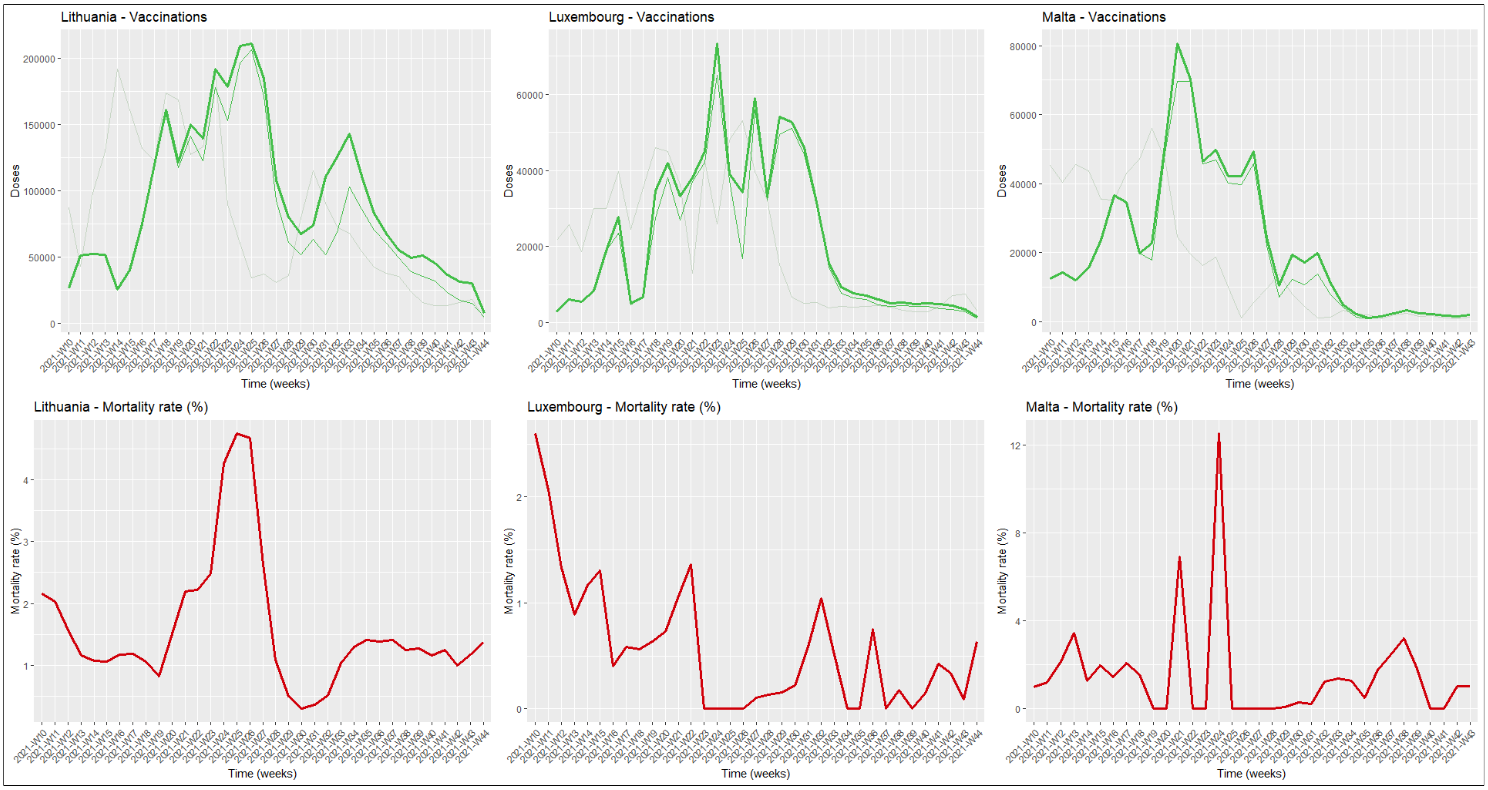

First Doses
Second doses
Last doses
(second + single)
Mortality rate

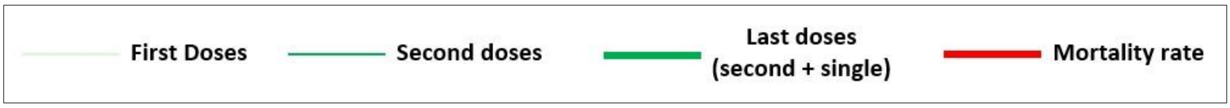

France - Vaccinations
Doses
Time (weeks)
Germany - Vaccinations
Doses
Time (weeks)
Greece - Vaccinations
Doses
Time (weeks)
France - Mortality rate (%)
Mortality rate (%)
Time (weeks)
Germany - Mortality rate (%)
Mortality rate (%)
Time (weeks)
Greece - Mortality rate (%)
Mortality rate (%)
Time (weeks)


Hungary - Vaccinations
Doses
Time (weeks)
Iceland - Vaccinations
Doses
Time (weeks)
Ireland - Vaccinations
Doses
Time (weeks)
Hungary - Mortality rate (%)
Mortality rate (%)
Time (weeks)
Iceland - Mortality rate (%)
Mortality rate (%)
Time (weeks)
Ireland - Mortality rate (%)
Mortality rate (%)
Time (weeks)

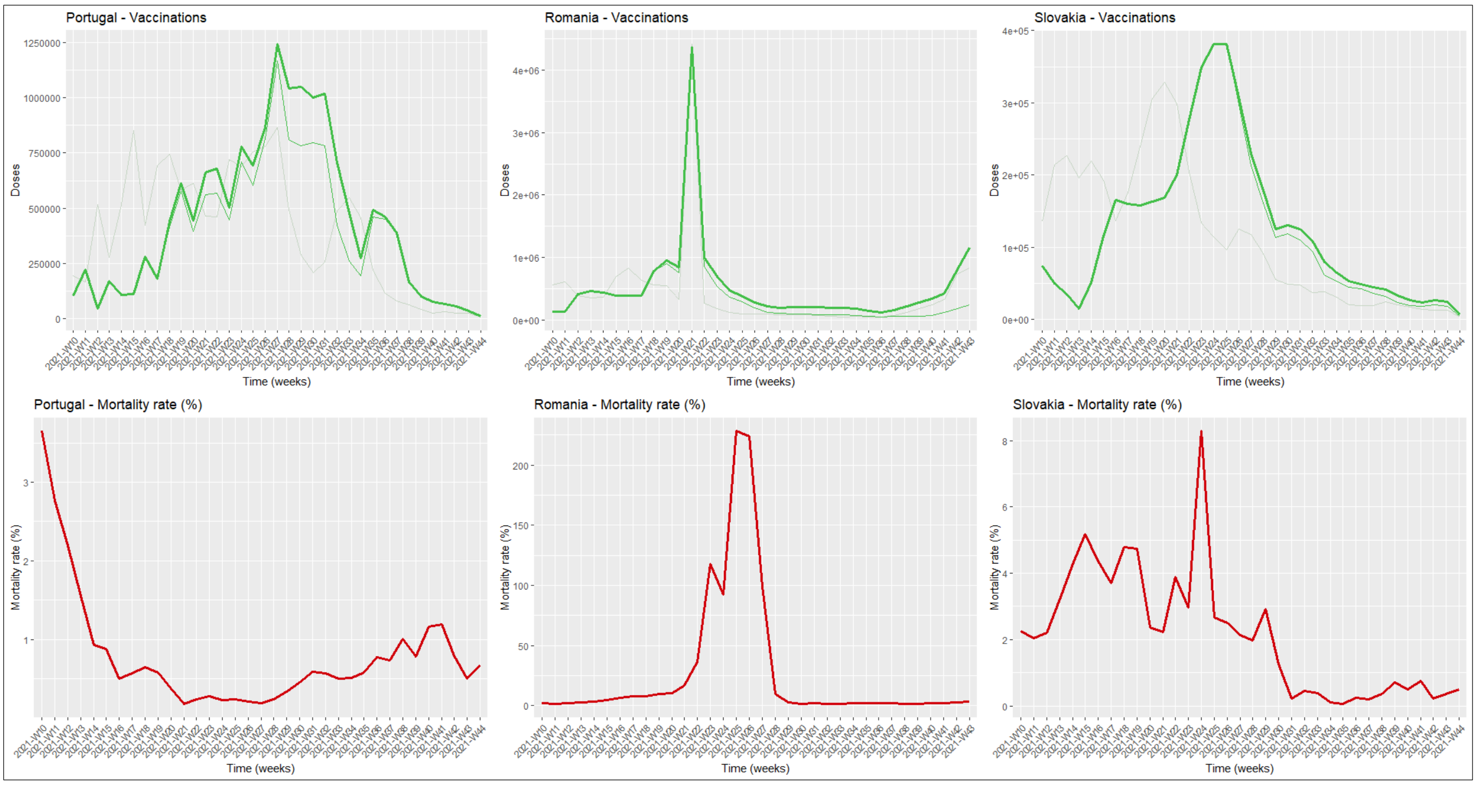

First Doses
Second doses
Last doses
(second + single)
Mortality rate


Italy - Vaccinations
Latvia - Vaccinations
Liechtenstein - Vaccinations
Doses
Time (weeks)
Italy - Mortality rate (%)
Latvia - Mortality rate (%)
Liechtenstein - Mortality rate (%)
Mortality rate (%)


Lithuania - Vaccinations
Luxembourg - Vaccinations
Malta - Vaccinations
Doses
Time (weeks)
Lithuania - Mortality rate (%)
Luxembourg - Mortality rate (%)
Malta - Mortality rate (%)
Mortality rate (%)


First Doses
Second doses
Last doses
(second + single)
Mortality rate


Netherlands - Vaccinations
Norway - Vaccinations
Poland - Vaccinations
Doses
Time (weeks)
Netherlands - Mortality rate (%)
Norway - Mortality rate (%)
Poland - Mortality rate (%)
Mortality rate (%)
Time (weeks)

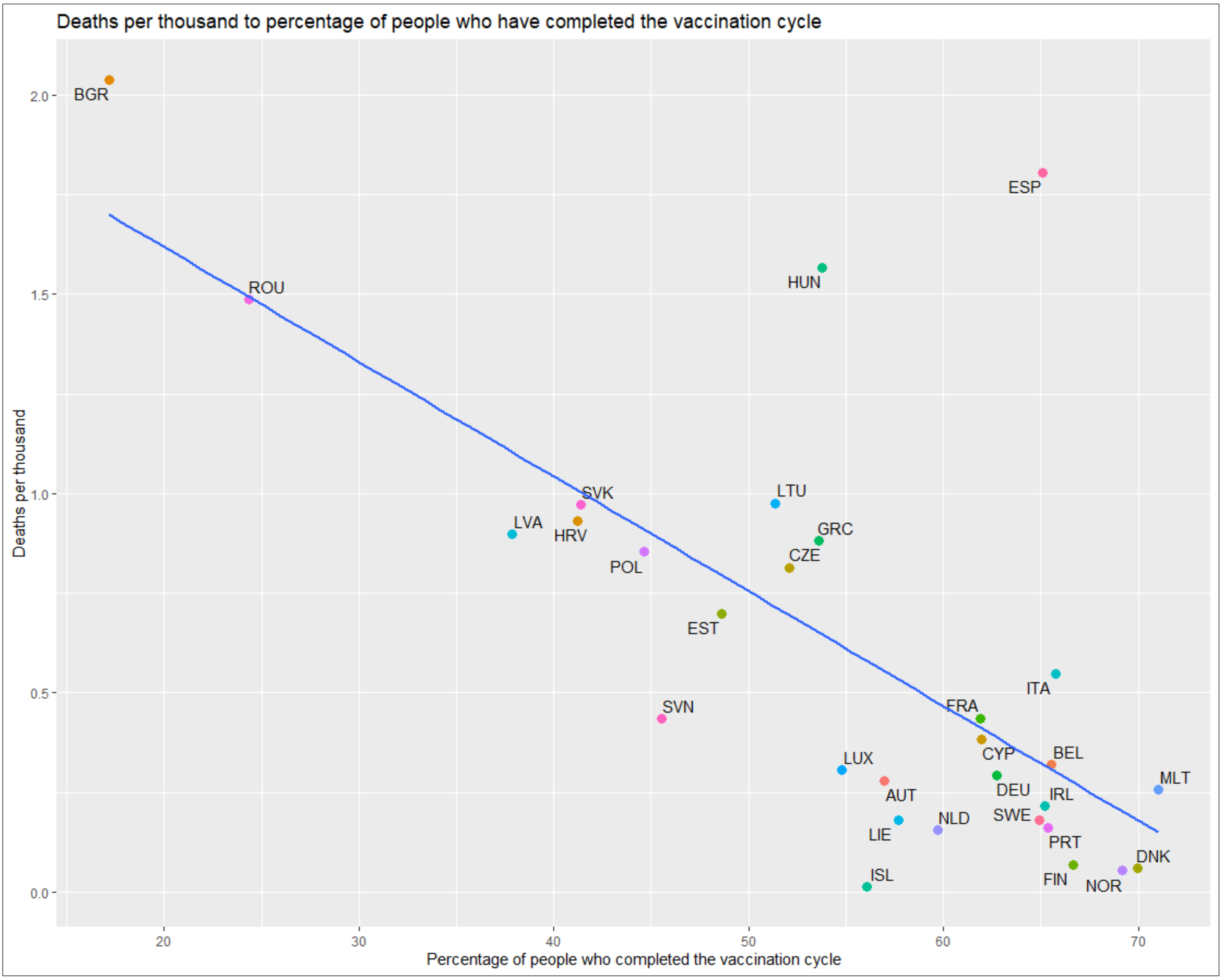

Portugal - Vaccinations
Romania - Vaccinations
Slovakia - Vaccinations
Doses
Time (weeks)
Portugal - Mortality rate (%)
Romania - Mortality rate (%)
Slovakia - Mortality rate (%)
Mortality rate (%)
Time (weeks)

There seems to be no clear correlation between the vaccination campaign and the mortality, but, in the following pages, I tried a different analysis to dig deeper into the possible correlation.

5.2 Deaths per thousand compared to the percentage of fully vaccinated people

The correlation I found, was between the deaths per thousand inhabitants of each country, and the percentage of people who completed the vaccination cycle.

In the scatter plot below, the correlation appears clearly, as the blue line, summing up the data thanks to the *geom_smooth(…)* function, is decreasing in deaths per thousand as the vaccination percentage grows.



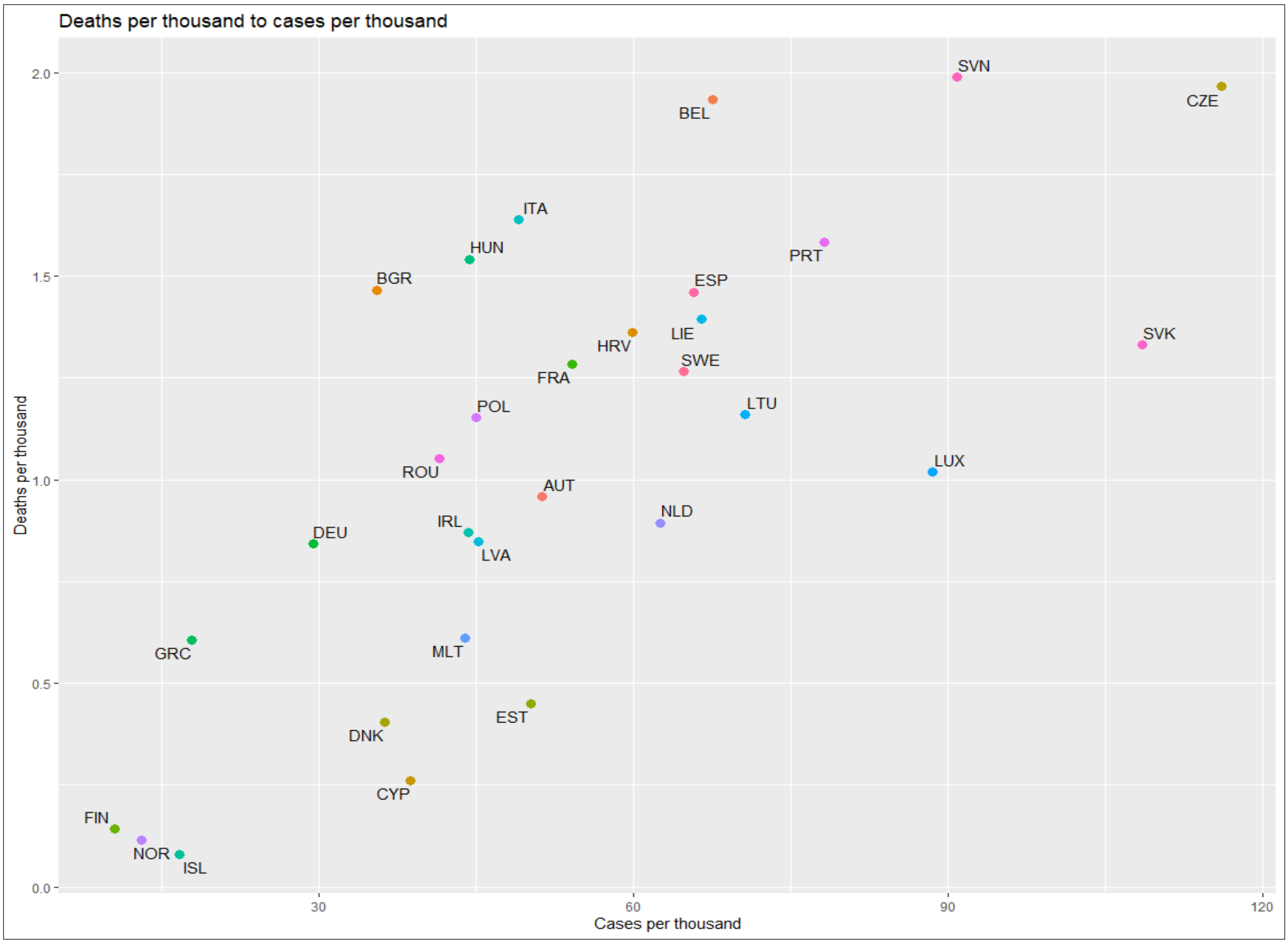
5.3 Deaths per thousand to cases per thousand evolutions before and after the vaccination campaign

A significant change in the cases to deaths correlation has taken place before and after the start of the vaccination campaign. The cases dataset starts on March 1st, containing in the first date all the data from the previous months. I separated the data after March 1st from the one contained in the March 1st row of every country.

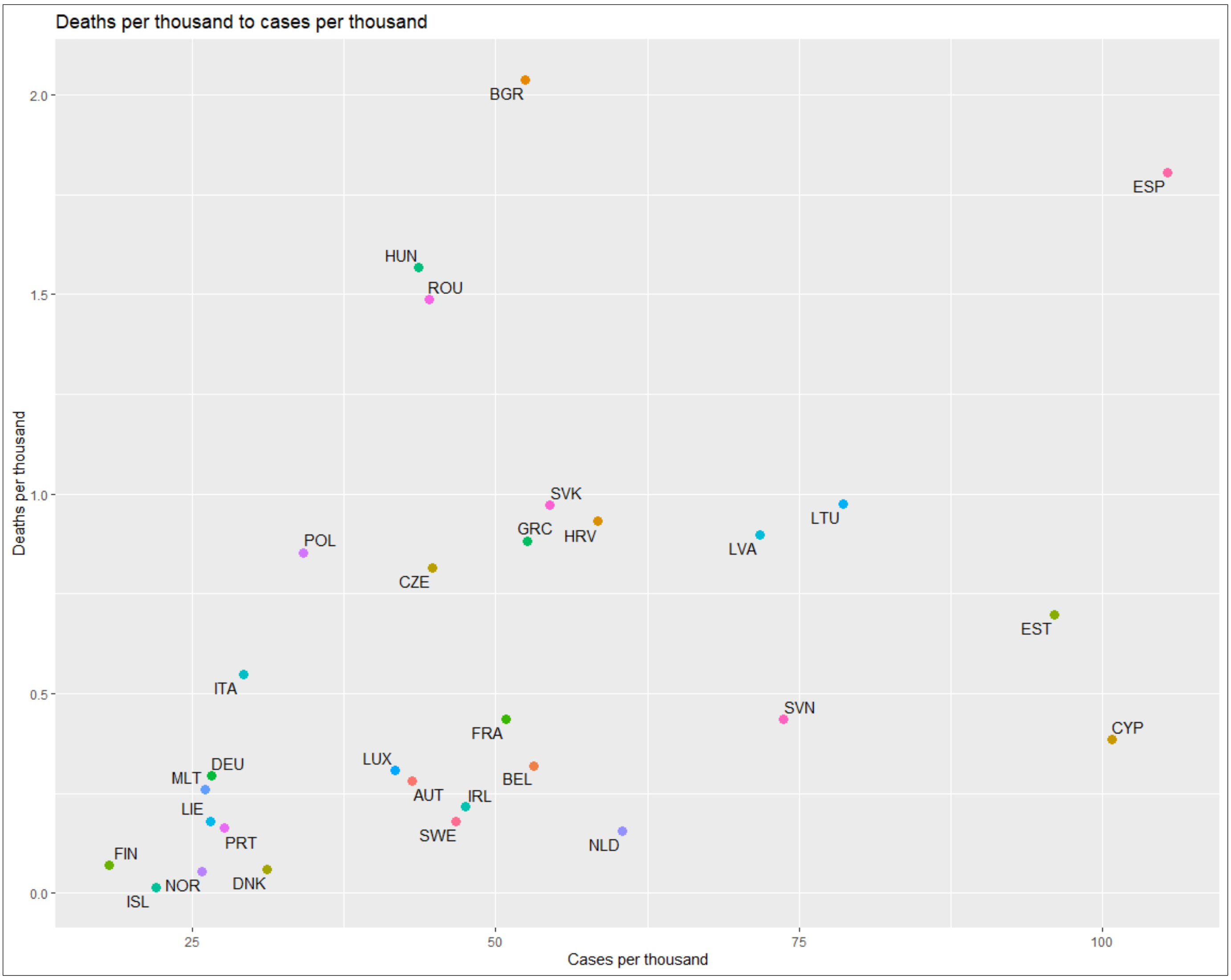
The vaccination campaign had already started in March, but the time frame it takes to become effective, and the fact that most vaccines were given later in the year, made me find it useful to make this comparison.

Here are the two scatter plots representing the correlation between deaths and cases pre and post vaccination campaign. The differences between the correlations, are well highlighted in the third graph, in which we see that there is a general decline in deaths, regardless of the evolution of the number of cases.

PRE-VACCINATION CAMPAIGN



POST-VACCINATION CAMPAIGN



PRE AND POST VACCINATION CAMPAIGN EVOLUTION