Covid-19 Exploratory Data Analysis using R

Theodoros Panagiotis Vagenas

PhD student

tpvagenas@mail.ntua.gr

Athens 2021

Introduction

The following analysis utilizes up-to-date Covid-19 data from John Hopkins CSSE containing information about cumulative confirmed cases and deaths on different countries during the interval 22/01/20 - 12/01/21. At first, the dataset is preprocessed in order to be analysed. Then summarized measurements and graphs are produced for public use. The graphs are drawn for the top 20 countries with the most Covid-19 cases worldwide. Finally, conclusions are extracted for the European Union countries including Greece,to which we have special interest. All the measurements and conclusions are accompanied with graphs (implemented codes in R provided).

1.1 R libraries

Below there is a list of the R libraries used in the codes presented in the project:

```
library(data.table)
library(date)
library(lubridate)
library(dplyr)
library(ggplot2)
library(cowplot)
library(gridExtra)
library(ggrepel)
library(arules)
library(ggExtra)
```

1.2 Data preparation

In this section, methods applied in order to clean and transform the data.

```
# Load Data for cumulative confirmed cases and deaths:
cases_data <- fread('time_series_covid19_confirmed_global.csv')
deaths_data <- fread('time_series_covid19_deaths_global.csv')
#The following columns are deleted from both datasets:
cases_data$'Province/State' <- NULL
cases_data$Lat <- NULL
cases_data$Long <- NULL
deaths_data$'Province/State' <- NULL
deaths_data$Lat <- NULL
deaths_data$Long <- NULL
#Data conversion from wide to long format:
cases.long <- melt(cases_data)
deaths.long <- melt(deaths_data)
#The produced columns are renamed to confirmed and deaths, and Country.Region to
    Country:
setnames(cases.long,"Country/Region","Country")
setnames(deaths.long,"Country/Region","Country")
setnames(cases.long,"value","confirmed")
setnames(deaths.long,"value","deaths")
#Conversion of date variable from character type to date object:
cases.long$variable <- cases.long[,lapply(.SD,mdy),.SDcols="variable"]
setnames(cases.long,"variable","date")
deaths.long$variable <- deaths.long[,lapply(.SD,mdy),.SDcols="variable"]
setnames(deaths.long,"variable","date")
#Group by country and date:
cases.long2 <- cases.long[,sum(confirmed),by = .(Country,date)]
setnames(cases.long2,"V1","confirmed")
deaths.long2 <- deaths.long[,sum(deaths),by = .(Country,date)]
setnames(deaths.long2,"V1","deaths")
#Merge the datasets:
com <- merge(cases.long2,deaths.long2,by=c("Country","date"))
#Calculate counts for the whole world: cases= 91595225, deaths=1962228
world <- com[,.(confirmed = sum(confirmed),deaths = sum(deaths)), by=date]
world$confirmed[length(world$confirmed)]
world$deaths[length(world$deaths)]
```

```r
#Sort by country and date:
com <- com[order(Country,date)]
#Calculate daily confirmed cases and deaths:
day1 <-min(com$date)
com <- com %>% mutate(confirmed.inc = ifelse(date==day1, NA, confirmed- lag(
    confirmed, n=1)))
com <- com %>% mutate(deaths.inc = ifelse(date==day1, NA, deaths- lag(deaths, n=1)))
```

Exploratory Data Analysis

2.1 World Covid-19 cases and deaths Analysis

First, a table with the same columns as table ”com” is created with com data grouped by date with summation. As a result, confirmed cases and deaths are summed up for each date to compute worldwide results.

```r
max_date <- max(com$date)
world_data_per_day <- com[,.(cases=sum(confirmed),deaths=sum(deaths),world.confirmed
    .inc=sum(confirmed.inc),world.deaths.inc=sum(deaths.inc)),by=date]
# Set English as language for date labels
Sys.setlocale("LC_TIME", "C")
```

In Figure 2.1 we can see the cumulative cases and deaths as they are represented graphically for every date worldwide. Confirmed cases escalated very fast. There is a period after November that the number of cases and deaths have a quite stable increase. As it was expected there is a strong correlation between the number of deaths and the number of cases.

```r
g1 <- ggplot()+
  geom_line(data=world_data_per_day,aes(x=date,y=cases),color="Blue") +
  labs(x = "Date", y="Cases",title=paste0("Worldwide cases - ",max_date))+
  scale_x_date(date_breaks="1 month",date_labels = "%b")+
  theme(legend.title=element_blank(),
        legend.position='bottom',
        plot.title =element_text(size = 13,hjust = 0.5))
g2 <- ggplot()+
  geom_line(data=world_data_per_day,aes(x=date,y=deaths),color="Blue")+
  labs(x = "Date", y="Deaths",title=paste0("Worldwide deaths - ",max_date))+
  scale_x_date(date_breaks="1 month",date_labels = "%b")+
  theme(legend.title=element_blank(),
        legend.position='bottom',
        plot.title = element_text(size = 13,hjust = 0.5))
plot_grid(g1, g2, labels=c("", "", ""), ncol = 2, nrow = 1, align = "v")
```

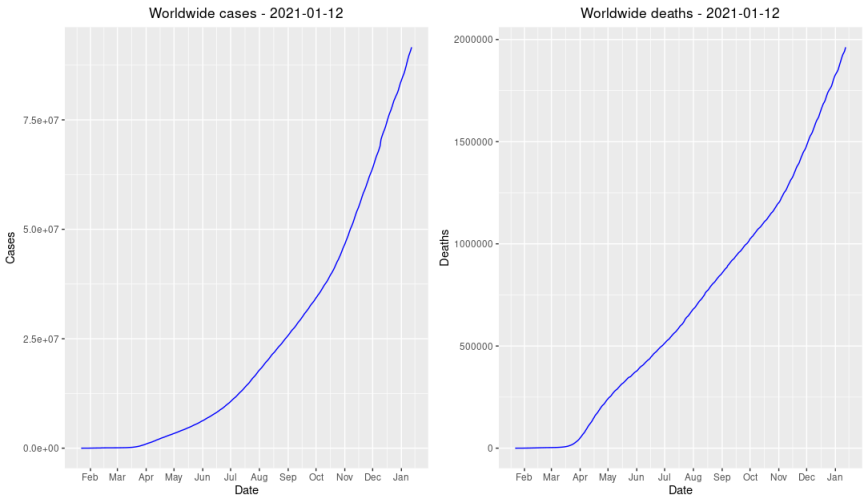

Figure 2.1: Worldwide cases and deaths

In Figure 2.2 the same results are plotted together in the logarithmic scale. This way a number of values can be displayed without compressing down the smallest values of cases and deaths. Also with logarithmic

scale the constant percentage of the increase in the cumulative cases/deaths is shown as a constant vertical increase. So, a straight line means a constant growth rate. As a result, we can observe two large sudden increases in growth rate the first months and a more smooth continuously increasing growth rate after May.

```
g2 <- ggplot()+
  geom_line(data=world_data_per_day,aes(x=date,y=cases,color='cases')) +
  geom_line(data=world_data_per_day,aes(x=date,y=deaths,color='deaths'))+
  labs(x = "Date", y="Cases",title="Worldwide number of cases and deaths in log
    scale")+
  theme(legend.title=element_blank(),
        legend.position='right',
        plot.title =element_text(hjust=0.5))+
  scale_y_continuous(trans='log10')+
  scale_x_date(date_breaks="1 month",date_labels = "%b")
plot(g2)
```

Figure 2.2: Worldwide cases and deaths (log scale)

In Figure 2.3 we can see the daily confirmed cases and deaths for the whole world for the same interval. Confirmed cases have some spikes between periods of smaller increases in new cases per day. In April, August and November these spikes are bigger so the everyday cases became suddenly much larger. Daily deaths also shown a big increase throughout April-May and also in November keeping a stable course in-between the two periods. After November daily new cases and deaths have been increased from day to day indicating an outbreak.

```
g3 <- ggplot(data=world_data_per_day,aes(x=date,y=world.confirmed.inc),color='blue')
    +
  geom_point() +
  geom_smooth()+
  labs(x = "Date", y="Cases",title="Worldwide daily cases")+
  scale_x_date(date_breaks="1 month",date_labels = "%b")+
  theme(plot.title =element_text(hjust=0.5))
g4 <- ggplot(data=world_data_per_day,aes(x=date,y=world.deaths.inc),color='blue')+
  geom_point() +
  geom_smooth()+
  labs(x = "Date", y="Cases",title="Worldwide daily deaths")+
  scale_x_date(date_breaks="1 month",date_labels = "%b")+
  theme(plot.title =element_text(hjust=0.5))
grid.arrange(g3, g4, ncol=2)
```

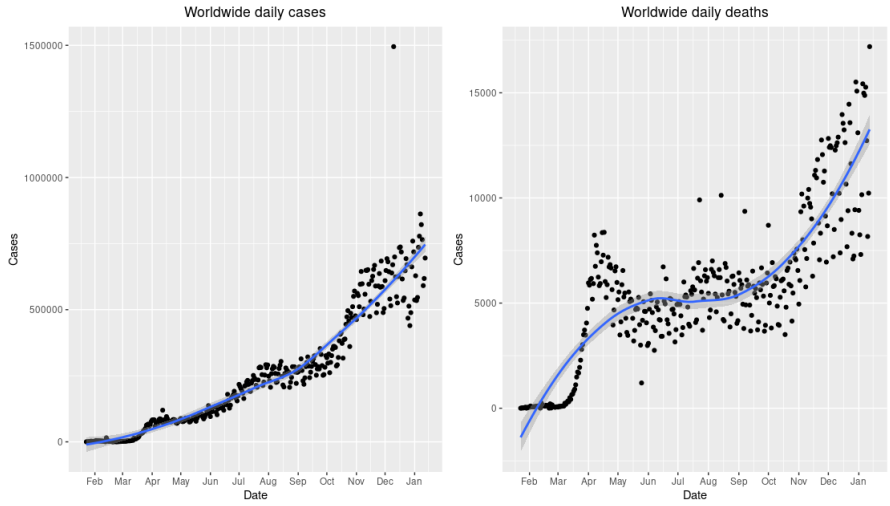

Figure 2.3: Worldwide cases and deaths

In the next step a death rate is calculated from the cumulative confirmed cases and deaths with the following equation:

$$death\_rate = 100\frac{deaths}{cases} \tag{2.1}$$

The code for this calculation and the plot is provided below:

```
world_data_per_day <- world_data_per_day %>% mutate(death_rate=100*deaths/cases)
g5 <- ggplot()+
  geom_line(data=world_data_per_day,aes(x=date,y=death_rate)) +
  labs(x = "Date", y="Death rate",title="Worldwide death rate")+
  scale_x_date(date_breaks="1 month",date_labels = "%b")+
  theme(legend.title=element_blank(),
        legend.position='bottom',
        plot.title = element_text(hjust=0.5))
plot(g5)
ggsave("death_rate.jpg", dpi=300)
```

In Figure 2.4 we can observe that the biggest value of death rate was between April and May which was the period after the first outbreak. The next months there was a decreasing tendency probably due to the lock-downs and restrictions that were imposed and due to the better understanding and treatment of the Covid-19.

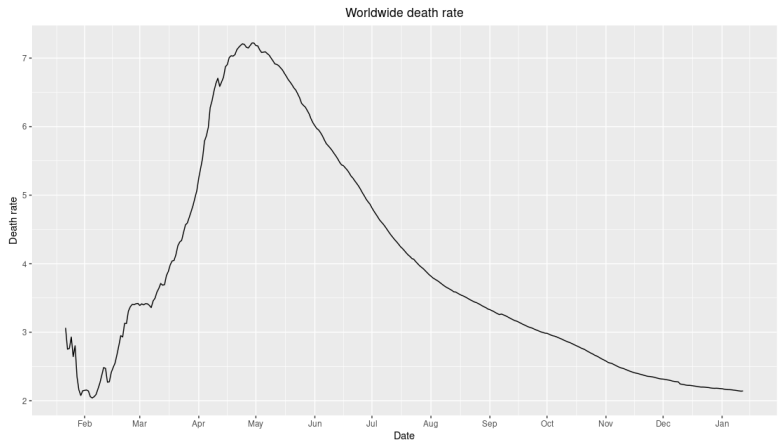

Figure 2.4: Worldwide death rate

2.2 Top 20 countries with the most confirmed cases

In this sections measurements and graphs are presented for the 20 countries with the most confirmed cases in the world. The countries are sorted by the number of cumulative cases. Also a column named ”other” is constructed and contains the cases and deaths for the rest countries. We can see that US have more Covid-19 cases and deaths than the rest of the countries, minus the top presented in the graph. In the second row, deaths are presented with the same order as before, so it can be easily observed that there are countries with more cases but less deaths than others with less cases. For example, Mexico is 13th in

confirmed cases but 4th in deaths from the top 20. This statistics may indicate inaccuracies in confirmed cases/deaths or a health system and generally a mechanism with smaller capability of dealing with the pandemic. This is more clear in the 3rd graph presenting the death rate, in which we can observe better the variability in deaths and death rate between countries with most cases.

```
top_countries <- com%>% filter(date== max(date))
top_countries$confirmed.inc <- NULL
top_countries$deaths.inc <- NULL
top_countries$date <- NULL
top_countries <- top_countries[order(-confirmed)]
top_countries <- top_countries %>% mutate(death_rate=100*deaths/confirmed)
other_countries <- top_countries[21:dim(top_countries)[1],lapply(.SD, sum),.SDcols =
    c("confirmed","deaths")]
other_countries_death_rate <- top_countries[21:dim(top_countries)[1],lapply(.SD,
   mean),.SDcols = c("death_rate")]
top_countries_20 <- rbind(top_countries[1:20,],cbind(Country="other",other_countries
   ,other_countries_death_rate))
top_countries_20$Country <- factor(top_countries_20$Country, levels = top_countries_
   20$Country)
```

```
# bar plot for comparison between confirmed cases, deaths and death rate
g6 <- ggplot(top_countries_20,aes(x = Country,y=confirmed,group=Country))+
  geom_bar(stat='identity')+
  geom_text(aes(label=confirmed, y=confirmed), vjust=0,size=2)+
  labs(x = element_blank(), y="Cases",title="Top 20 Countries with most cases: Cases
    ")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text(size =
    8))
g7 <- ggplot(top_countries_20,aes(x = Country,y=deaths,group=Country))+
  geom_bar(stat='identity')+
  geom_text(aes(label=deaths, y=deaths), vjust=0,size=2)+
  labs(x = element_blank(), y="Deaths",title="Top 20 Countries with most cases:
    Deaths")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text( size =
    8))
g8 <- ggplot(top_countries_20,aes(x = Country,y=death_rate,group=Country))+
  geom_bar(stat='identity')+
  geom_text(aes(label=round(death_rate,2), y=death_rate), vjust=0,size=3)+
  labs(x = "Country", y="Death rate",title="Top 20 Countries with most cases: Death
    rate")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text(size =
    8))
#grid.arrange(g6, g7,g8, nrow=3,top="Top 20 countries with the most confirmed cases
    ")
plot_grid(g6, g7, g8, labels=c("", "", ""), ncol = 1, nrow = 3, align = "v")
```

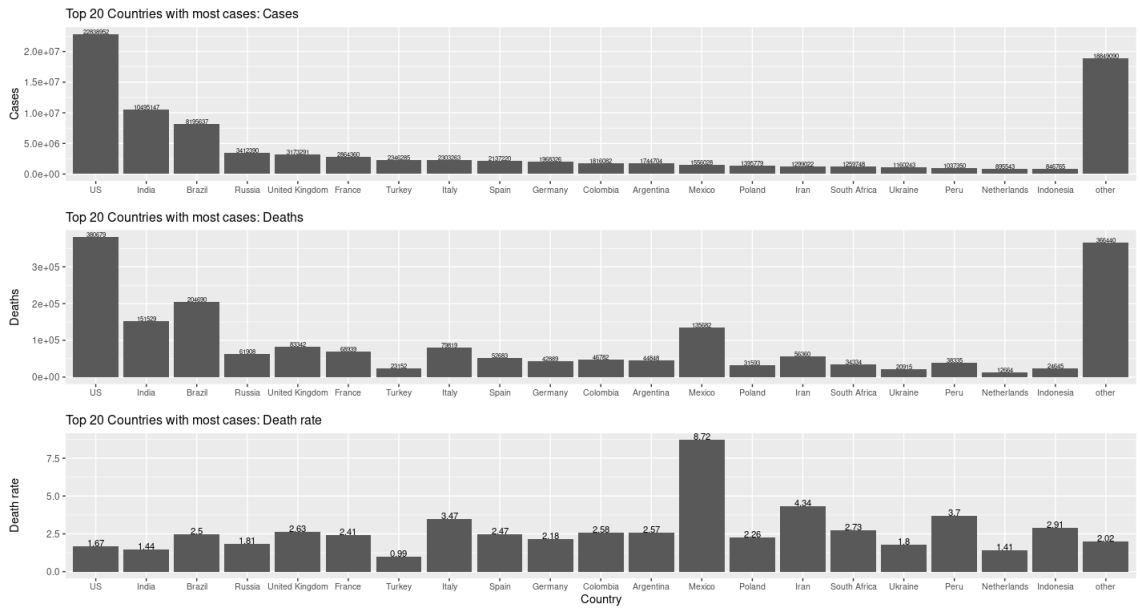

Figure 2.5: Top 20 countries cases, deaths and death rate

In Figure 2.6 cases are plotted as opposed to deaths for the previous 20 countries. Colors show the countries and the diameter of the points shows the death rate. This display depicts clearly the relationship between cases and deaths in these countries. Especially, there is a concentration of countries below 5 millions cases and 100 thousands deaths and two other distinctive smaller clusters that differ a lot from this concentration.

```
g9 <-  ggplot(data=top_countries_20,aes(x=confirmed,y=deaths,group=Country))+
  geom_point(aes(color=Country,size=death_rate)) +
  labs(x = "Confirmed cases", y="Deaths",title="Top 20 countries with most cases:
    Deaths vs confirmed",color="Country",size="Death rate")+
  geom_text_repel(aes(label = Country))+
  theme(legend.title=element_text(),legend.position='bottom',plot.title =element_
    text(hjust = 0.5,size=12),axis.text.x=element_text(size=10),axis.text.y=element_
    text(size = 10))
plot(g9)
ggsave("top20_confirmed_deaths.jpg", dpi=300)
```

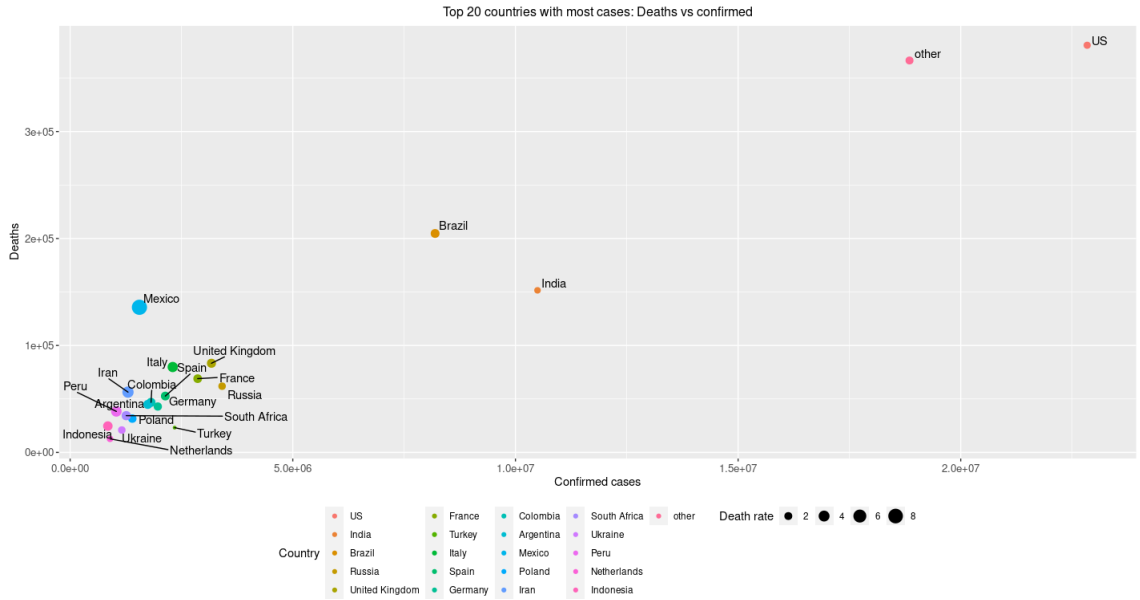

Figure 2.6: Top 20 Confirmed cases vs deaths

World map of cases

In the Figure 2.7 below, a map with confirmed cases is plotted. According to the color scaled graph, yellow indicates the countries with less cases and red with more cases. Also grey level is used for countries that are not in the dataset. There was also some differences in region names that were corrected manually in order to generate graph for the most of the countries. In the map, we can see the spatial distribution of the Covid-19 cases until the last date of the dataset.

```
# Construct world map with confirmed cases
# Load again dataset without dropping columns
full_world <- map_data("world")
cases_data <- fread('time_series_covid19_confirmed_global.csv')
deaths_data <- fread('time_series_covid19_deaths_global.csv')
cols_vector = c('Province/State','Country/Region','Lat','Long',names(cases_data)[
    length(names(cases_data))])
cases_data <- cases_data[,..cols_vector]
cases_data$'Province/State' <- NULL
setnames(cases_data,"Country/Region","region")
cases_data <- cases_data[,.(lat=Lat,long=Long,confirmed=sum(`1/12/21`)),by=.(region)
    ] # here last date
sequence_break <- c(1,100,1000,10000,100000,1000000)
full_world$region[full_world$region=="USA"] <- "US"
full_world$region[full_world$region=="Myanmar"] <- "Burma"
full_world$region[full_world$region=="Macedonia"] <- "North Macedonia"
full_world$region[full_world$region=="South Korea"] <- "Korea, South"
full_world$region[full_world$region=="UK"] <-"United Kingdom"
full_world$region[full_world$region=="Republic of Congo"] <- "Congo (Brazzaville)"
```

```
full_world$region[full_world$region=="Democratic Republic of the Congo"] <- "Congo (
    Kinshasa)"
full_world$region[full_world$region=="Ivory Coast"] <- "Cote d'Ivoire"
full_world$region[full_world$region=="Taiwan"] <- "Taiwan*"
full_world$region[full_world$region=="Czech Republic"] <- "Czechia"
full_world$region[full_world$region=="French Guiana"] <- "Czechia"
nn <- left_join(cases_data, full_world, by = "region")
gworldmap <- ggplot()+
  geom_polygon(data = full_world, aes(x=long,y=lat,group=group),alpha=0.2)+
  geom_polygon(data = nn, aes(x=long.y,y=lat.y,group=group,fill=confirmed),alpha=1)+
  scale_fill_gradientn(colours = rev(heat.colors(7)),breaks = sequence_break,trans =
     "log10")+
  annotate(geom="text", x=150, y=-100, label="Grey countries' measurements are not
    available",color="red",alpha=0.7)+
  theme(
    plot.title =element_text(hjust = 0.5),
    legend.position = "right",
    legend.background = element_rect(fill = "#ffffff", color = NA)
  )
plot(gworldmap)
```

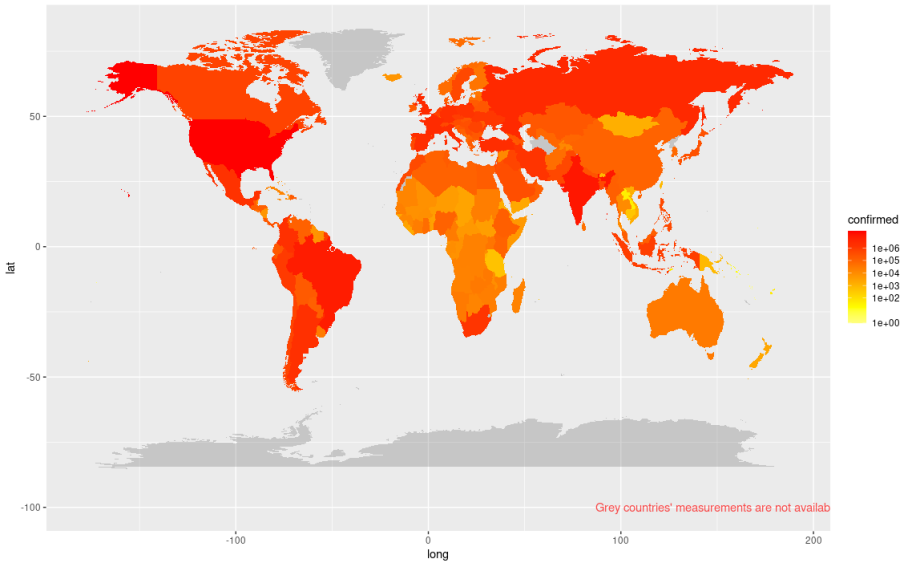

Figure 2.7: World map with Covid-19 confirmed cases

2.3 European Union countries analysis

The next region of interest is the European Union countries with a special analysis for Greece. It can also provide more information about the effects and the Covid-19 progression between neighbor countries . The appropriate table is constructed similar to the ”Top 20 countries” case.

```
europeanUnion <- c("Austria","Belgium","Bulgaria","Croatia","Cyprus",
                   "Czech Rep.","Denmark","Estonia","Finland","France",
                   "Germany","Greece","Hungary","Ireland","Italy","Latvia",
                   "Lithuania","Luxembourg","Malta","Netherlands","Poland",
                   "Portugal","Romania","Slovakia","Slovenia","Spain",
                   "Sweden","United Kingdom")
europe_com <- com %>% filter(com$Country %in% europeanUnion)
euro_countries <- europe_com%>% filter(date== max(date))
euro_countries$confirmed.inc <- NULL
euro_countries$deaths.inc <- NULL
euro_countries$date <- NULL
euro_countries <- euro_countries[order(-confirmed)]
euro_countries <- euro_countries %>% mutate(death_rate=100*deaths/confirmed)
euro_countries$Country <- factor(euro_countries$Country, levels = euro_countries$
    Country)
temp <- europe_com %>% filter(Country %in% euro_countries$Country)
euro_countries <- euro_countries %>% mutate(categ = discretize(euro_countries$death_
    rate, method = "frequency", breaks = 3))
# European confirmed cases,deaths and death rates ordered by cases (descent)
g18 <- ggplot(euro_countries,aes(x = Country,y=confirmed,group=Country))+
```

```
  geom_bar(stat='identity')+
  geom_text(aes(label=round(confirmed, 2), y=confirmed), size=2, vjust=0)+
  labs(x = "", y="Cases",title="Europe: Confirmed cases")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text(size =
    7))
g19 <- ggplot(euro_countries,aes(x = Country,y=deaths,group=Country))+
  geom_bar(stat='identity')+
  geom_text(aes(label=round(deaths, 2), y=deaths), size=2, vjust=0)+
  labs(x = "", y="Deaths",title="Europe: Deaths")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text(size =
    7))
g20 <- ggplot(euro_countries,aes(x = Country,y=death_rate,group=Country))+
  geom_bar(stat='identity')+
  geom_text(aes(label=round(death_rate, 2), y=death_rate), size=2, vjust=0)+
  labs(x = "Country", y="Death rate",title="Europe: Death rate")+
  theme(plot.title =element_text(hjust = 0,size=12),axis.text.x=element_text(size =
    7))
# align vertical for more clear results
plot_grid(g18, g19, g20, labels=c("", "", ""), ncol = 1, nrow = 3, align = "v")
```

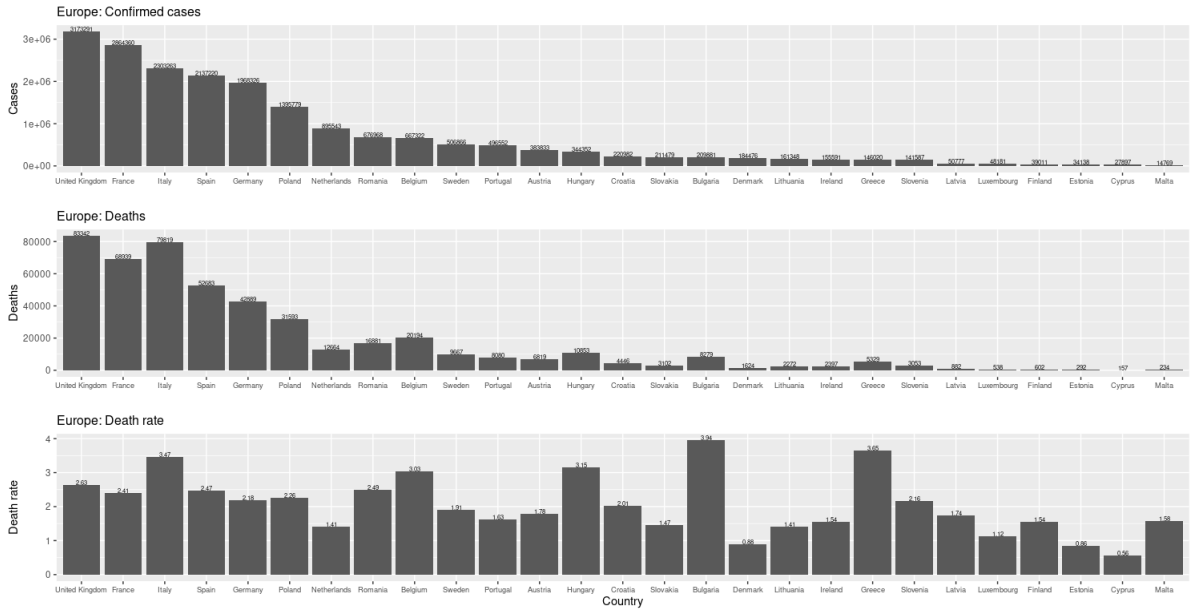

Figure 2.8: European Union countries cases, deaths and death rate

In Figure 2.8 above European Union members are sorted by confirmed cases. There is a vertical alignment so they can be compared to each other. An interesting point of view the study of the reasons that lead to less cases in them so that can help the other countries to protect themselves from another outbreak. Death rate can also be an indication that some countries confronted Covid-19 better. Biggest death rate here is observed in Italy and UK where there was a very fast increase in cases, a situation that caused problems in the health system. Death rate also has a different behavior from cases as we can see from the fact that countries are sorted by the cases.

The next diagram Figure 2.9 plots deaths versus cases. The form of the graph between these two variables for these countries is very similar but has different scales. This fact was also observed in the previous plots that also show the death rates.

```
g21 <-  ggplot(data=temp,aes(x=confirmed,y=deaths,group=Country))+
  geom_line(aes(color=Country)) +
  labs(x = "Confirmed cases", y="Deaths",title="Europe: Deaths vs Confirmed cases")+
  theme(legend.title=element_blank(), legend.position='right',plot.title =element_
    text(hjust = 0.5))
plot(g21)


g22 <-  ggplot(data=euro_countries,aes(x=confirmed,y=deaths,group=Country))+
  geom_point(aes(color=Country,size=death_rate)) +
  labs(x = "Confirmed cases", y="Deaths",title="Europe: Deaths vs Confirmed cases",
    color="Country",size="Death rate")+
```

```
  geom_text_repel(aes(label = Country))+
  theme(legend.title=element_text(),legend.position='bottom',plot.title =element_
    text(hjust = 0.5,size=12),axis.text.x=element_text(size=10),axis.text.y=element_
    text(size = 10))
plot(g22)
```

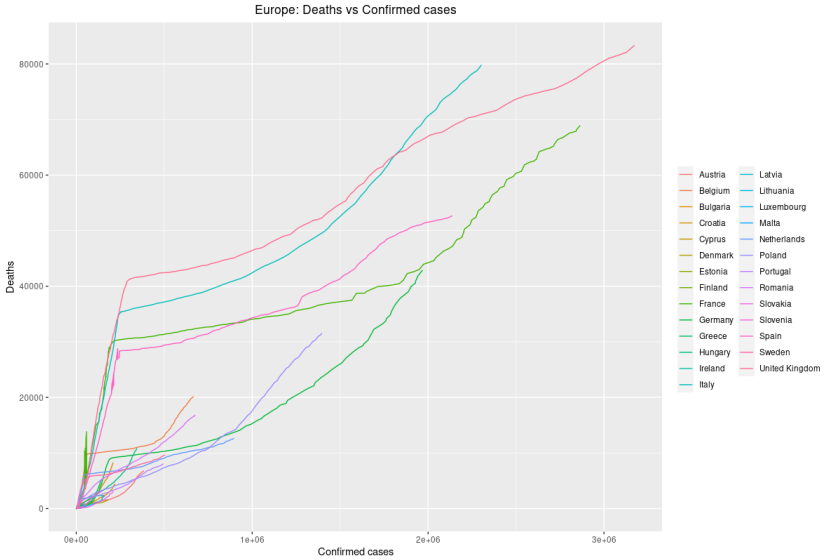

Figure 2.9: European Union countries deaths vs confirmed

In the next diagram Figure 2.10 the same plot is presented but with points and death rates, in correspondence with the ”Top 20 countries” occasion.

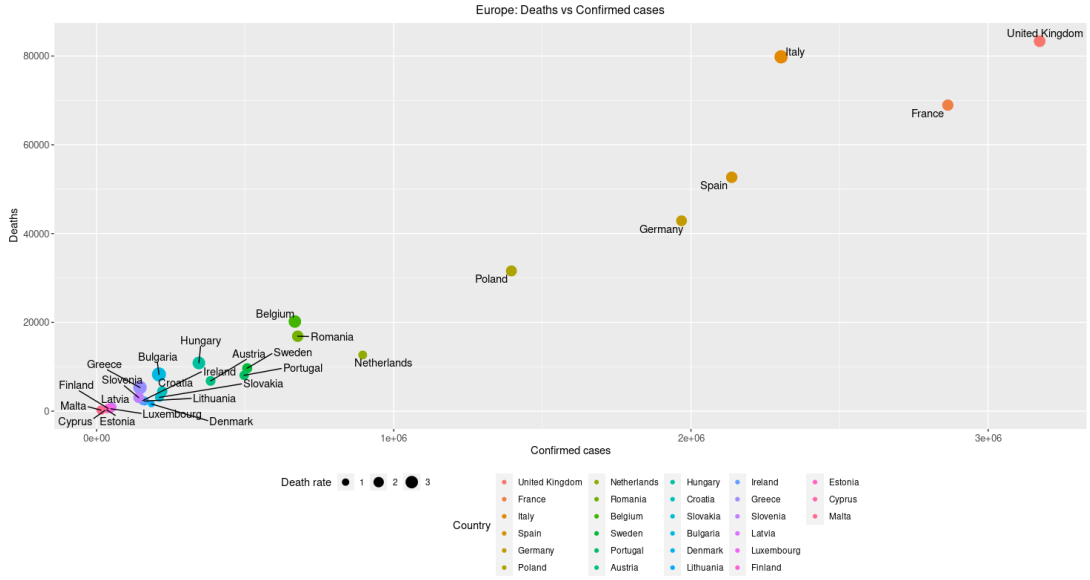

Figure 2.10: European Union countries deaths vs confirmed

With the diverging bars below (Figure 2.11) we can observe the deviations between EU countries according to cases and deaths. It shows how much the countries differ from the average. So we can see that countries with cases and deaths below the average are more close to the average/midpoint. This fact indicates that Covid-19 crisis is lower in the most EU countries. On the contrary, countries above the average are further than midpoint, which is an indication that these countries have a much difficult situation than the others.

```

# Diverging bars Europe confirmed and deaths
euro_countries$confirmed_norm <- round((euro_countries$confirmed - mean(euro_
    countries$confirmed))/sd(euro_countries$confirmed), 2)
euro_countries$confirmed_type <- ifelse(euro_countries$confirmed_norm < 0, "below","
    above")
euro_countries <- euro_countries[order(euro_countries$confirmed_norm), ] # sort
```

```
g25 <- ggplot(euro_countries, aes(x=Country, y=confirmed_norm, label=confirmed_norm)
    ) +
  geom_bar(stat='identity', aes(fill=confirmed_type), width=.5) +
  scale_fill_manual(name="Confirmed cases",labels = c("Above Average", "Below
    Average"),
                    values = c("above"="#00ba38", "below"="#f8766d")) +
  labs(subtitle="Normalised confirmed cases",
       title= "Diverging Bars for Europe") +
  theme( legend.position='bottom')+
  coord_flip()

euro_countries$deaths_norm <- round((euro_countries$deaths - mean(euro_countries$
    deaths))/sd(euro_countries$deaths), 2)
euro_countries$deaths_type <- ifelse(euro_countries$deaths_norm < 0, "below","above"
    )
euro_countries <- euro_countries[order(euro_countries$deaths_norm), ] # sort
g26 <- ggplot(euro_countries, aes(x=Country, y=deaths_norm, label=deaths_norm)) +
  geom_bar(stat='identity', aes(fill=deaths_type), width=.5) +
  scale_fill_manual(name="Deaths",labels = c("Above Average", "Below Average"),
                    values = c("above"="#00ba38", "below"="#f8766d")) +
  labs(subtitle="Normalised deaths",
       title= "Diverging Bars for Europe") +
  theme( legend.position='bottom')+
  coord_flip()
grid.arrange(g25,g26, ncol=2)
```

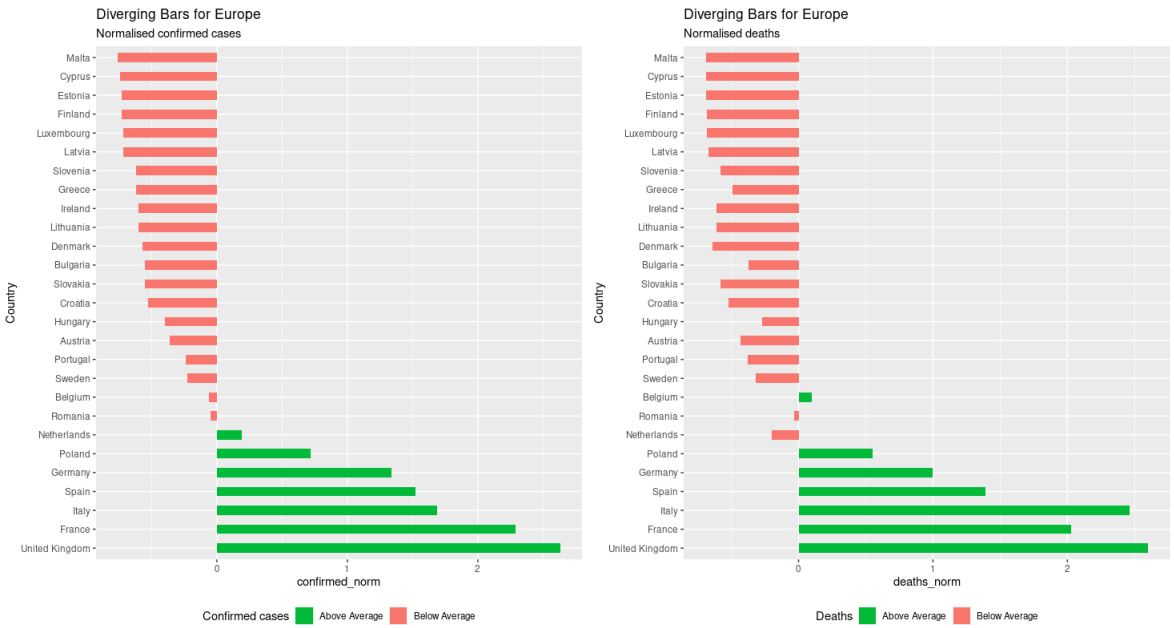

Figure 2.11: Diverging bars for European Union deaths and confirmed cases

In Dot plot Figure 2.12 distances between EU countries according to cases and deaths are better visualized. As it was already mentioned there is one large group of countries with small number of cases and deaths and some countries with larger numbers. This diagram emphasizes the fact that countries with many cases/deaths present numbers significant larger than countries with less cases/deaths. Greece is in the countries with very few cases as opposed to the others.

```
# Ranking for Europe
g27 <- ggplot(euro_countries, aes(y=Country, x=confirmed)) +
  geom_point(col="tomato2", size=3) + # Draw points
  geom_segment(aes(y=Country,
                   yend=Country,
                   x=min(confirmed),
                   xend=max(confirmed)),
                   linetype="dashed",
                   size=0.1)+
  labs(title="Dot Plot for Europe",subtitle="Country vs Confirmed")

g28 <- ggplot(euro_countries, aes(y=Country, x=deaths)) +
```

```
  geom_point(col="tomato2", size=3) + # Draw points
  geom_segment(aes(y=Country,
                   yend=Country,
                   x=min(deaths),
                   xend=max(deaths)),
                   linetype="dashed",
                   size=0.1)+
  labs(title="Dot Plot for Europe",subtitle="Country vs Deaths")
grid.arrange(g27,g28, ncol=2)
```

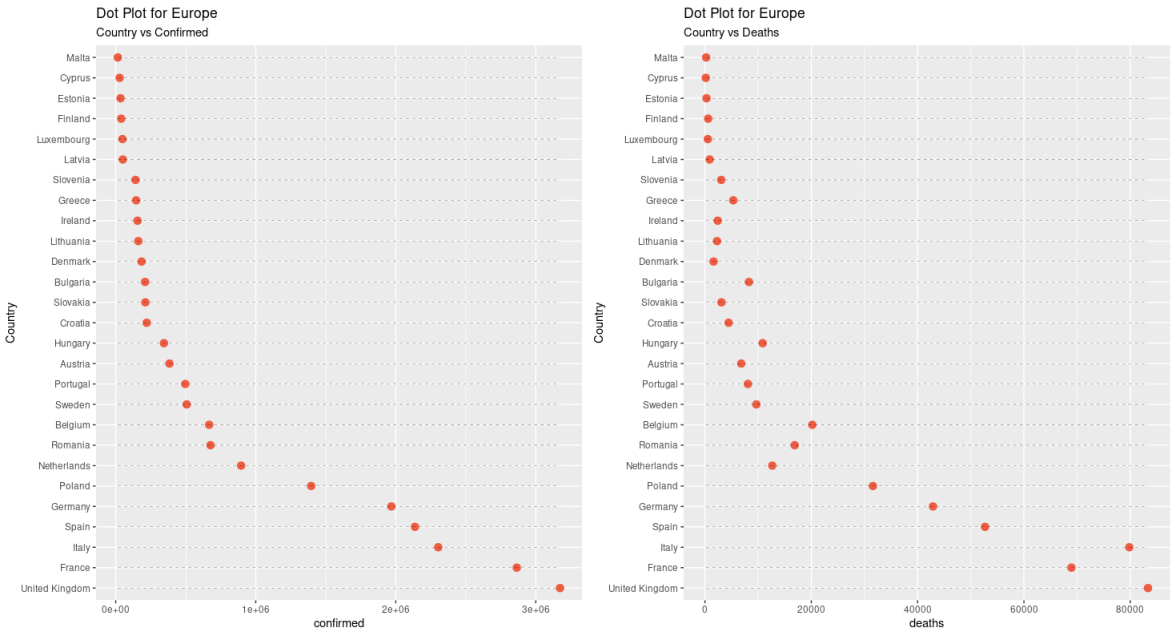

Figure 2.12: Ranking for European Union deaths and confirmed cases

2.4 Greece Covid-19 Analysis

In this section, graphs show the evolution of the Covid-19 pandemic in Greece. The results can underline the spread of confirmed cases during the past months and useful observations for the analysis of the situation in our country in comparison with the rest of the European Union countries. Daily cases are also calculated in order to get a better knowledge of the speed of the spreading. Also, conclusions about the distribution of daily cases and deaths are extracted.

In Figure 2.13 an exponential increase of cases and deaths is presented. From this plot and also from the Figure 2.14 we can observe that the behavior of cases and deaths is quite similar but of course the number of deaths is much smaller than the number of cases.

```
greece <- com %>% filter(Country=="Greece")
g11 <- ggplot(greece, aes(x=date, y=confirmed)) + geom_bar(stat="identity", width
    =0.1) +
  theme_classic() +
  labs(title = "Covid-19 Confirmed Cases in Greece", x= "Date", y= "Confirmed cases"
    ) +
  theme(plot.title = element_text(hjust = 0.5))
g12 <- ggplot(greece, aes(x=date, y=deaths)) + geom_bar(stat="identity", width=0.1)
    +
  theme_classic() +
  labs(title = "Covid-19 Deaths in Greece", x= "Date", y= "Deaths") +
  theme(plot.title = element_text(hjust = 0.5))
grid.arrange(g11, g12, ncol=2)
```

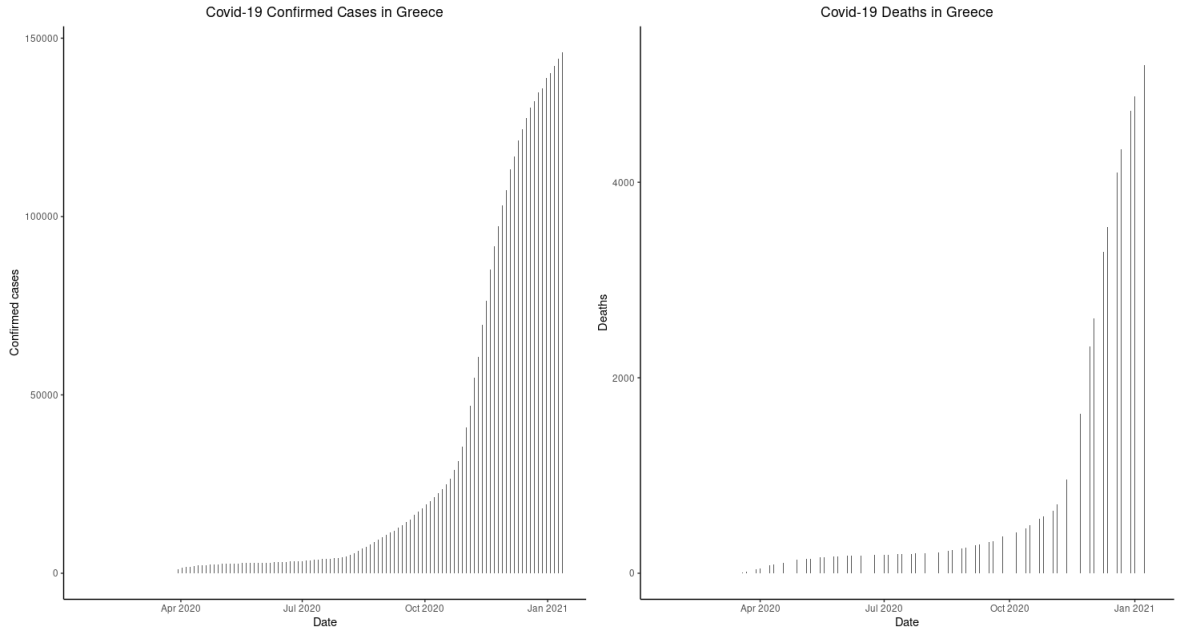

Figure 2.13: Greece confirmed cases and deaths versus date

```r
# Cases-deaths relationships
g13 <-  ggplot(data=greece,aes(x=confirmed,y=deaths))+
  geom_line() +
  labs(x = "Confirmed cases", y="Deaths",title="Greece: Confirmed cases vs Deaths")+
  theme(legend.title=element_blank(), legend.position='bottom',plot.title =element_
    text(hjust = 0.5))
plot(g13)
```

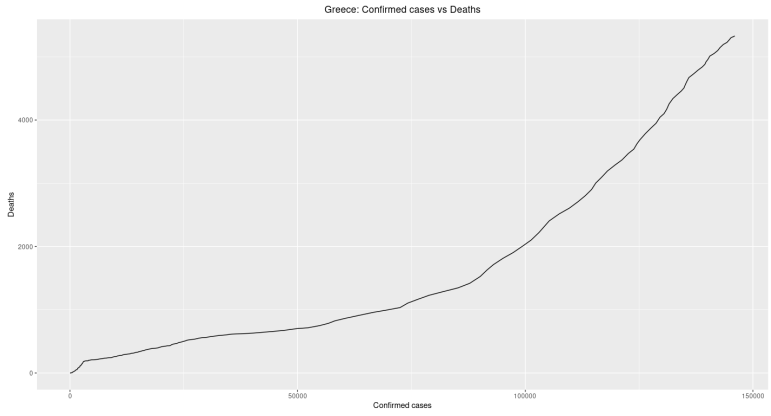

Figure 2.14: Greece confirmed cases versus deaths

In Figure 2.15 the new confirmed cases and deaths for each day are plotted. An initial increase is observed in September. After that in November there is a very big spike. This was the second outbreak in Greece when the second lockdown was imposed. In December the decrease is probably the result of the restrictions of the previous days.

```r
gg_daily1 <- ggplot(greece, aes(x=date, y=confirmed.inc)) + geom_line() +
  theme_classic() +
  labs(title = "Covid-19 Daily Confirmed Cases in Greece", x= "Date", y= "Confirmed
    cases") +
  theme(plot.title = element_text(hjust = 0.5))+
  scale_x_date(date_breaks="1 month",date_labels = "%b")
gg_daily2 <- ggplot(greece, aes(x=date, y=deaths.inc)) + geom_line() +
  theme_classic() +
  labs(title = "Covid-19 Daily Confirmed Deaths in Greece", x= "Date", y= "Confirmed
     deaths") +
  theme(plot.title = element_text(hjust = 0.5))+
  scale_x_date(date_breaks="1 month",date_labels = "%b")
grid.arrange(gg_daily1,gg_daily2, ncol=2)
```

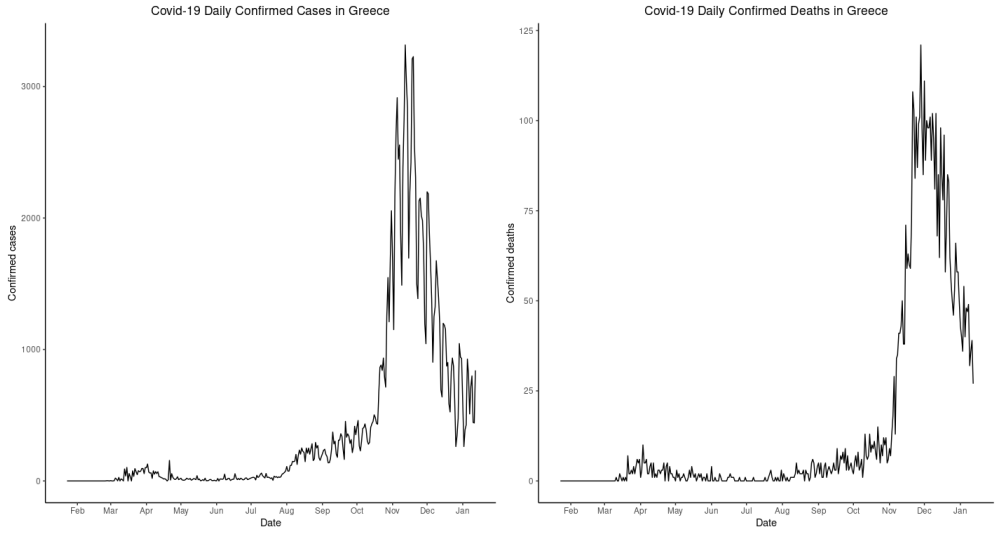

Figure 2.15: Greece: daily confirmed cases and deaths

In the following bar plot Figure 2.16 mean cases and deaths according to each day of week are presented. It is easily observed that Monday and Sunday are the days with the least confirmed cases probably due to the smaller number of tests that took place in the previous days (Saturday and Sunday accordingly). But this tendency is not followed by deaths that are not dependent to the tests conducted.

```
# Greece results per day of week
days_all <- greece %>% mutate(daysofweek=weekdays(date))
days_all <- na.omit(days_all)
days_all <- days_all[,.(confirmed = mean(confirmed.inc),deaths=mean(deaths.inc)),by=
    daysofweek]
days_all <- days_all %>% mutate(death_rate=100*deaths/confirmed)
days_all$daysofweek <- factor(days_all$daysofweek, levels = c("Monday","Tuesday","
    Wednesday","Thursday","Friday","Saturday","Sunday"))
g14 <- ggplot(days_all,aes(x = daysofweek,y=confirmed,fill=confirmed))+
  geom_bar(stat='identity')+
  geom_text(aes(label=round(confirmed, 2), y=confirmed), size=2, vjust=0)+
  labs(x = "Day", y="Confirmed cases",title="Greece: Mean of confirmed cases per day
     of week")+
  scale_fill_gradientn(colours = rev(heat.colors(7)))+
  theme(plot.title =element_text(hjust=0.5),axis.text.x=element_text(hjust=0.5))
g15 <- ggplot(days_all,aes(x = daysofweek,y=deaths,fill=deaths))+
  geom_bar(stat='identity')+
  geom_text(aes(label=round(deaths, 2), y=deaths), size=2, vjust=0)+
  labs(x = "Day", y="Deaths",title="Greece: Mean of deaths per day of week")+
  scale_fill_gradientn(colours = rev(heat.colors(7)))+
  theme(plot.title =element_text(hjust=0.5),axis.text.x=element_text(hjust=0.5))

grid.arrange(g14,g15, nrow=2)
```

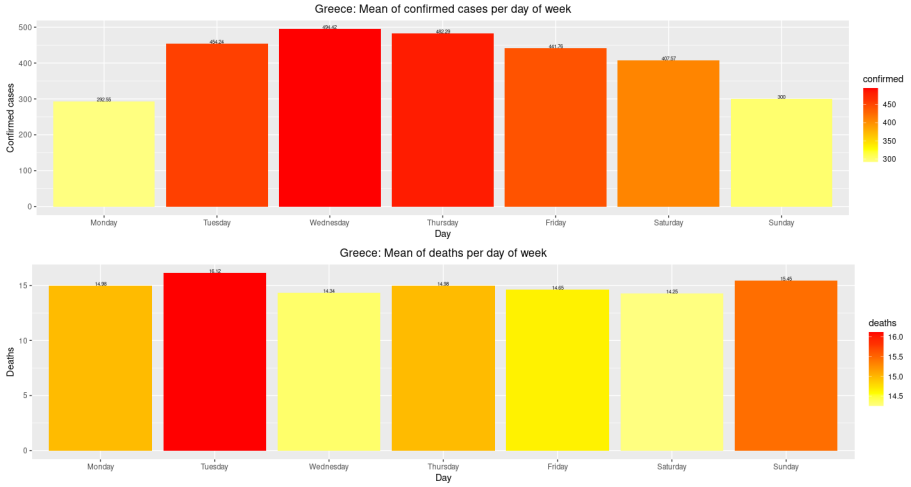

Figure 2.16: Greece confirmed cases and deaths versus day of week

Next there are two box plots Figure 2.17 for the distribution of the number of new cases and deaths in Greece every day. The median value for daily confirmed cases is somewhere lower than 100 and for deaths 0-5. This is affected by the fact that in the first outbreak of the pandemic Greece had taken restriction measurements immediately and for the first months the number of cases was very small. Also the interquartile range (IQR) for cases is 0-430 which is small according to last numbers of cases. Bigger numbers of cases/deaths are presented outside IQR because they are much less than the observations with high values due to the first part of the pandemic where Greece had small numbers of cases.

```
# Distribution for Greece
g29 <- ggplot(greece, aes(Country, confirmed.inc))+
  geom_boxplot(varwidth=T, fill="plum") +
  labs(title="Box plot for Greece",
       subtitle="Confirmed cases per day",
       x="",
       y="Confirmed cases")+
  background_grid()+
  theme(panel.grid.minor = element_line(colour="white", size=0.2))+
  scale_y_continuous(minor_breaks = seq(0 , 4000, 20), breaks = seq(0, 4000, 100))

g30 <- ggplot(greece, aes(Country, deaths.inc))+
  geom_boxplot(varwidth=T, fill="plum") +
  labs(title="Box plot for Greece",
       subtitle="Confirmed deaths per day",
       x="",
       y="Confirmed deaths")+
  background_grid()+
  scale_y_continuous(minor_breaks = seq(0 , 130, 5), breaks = seq(0, 130, 10))
grid.arrange(g29,g30, ncol=2)
```

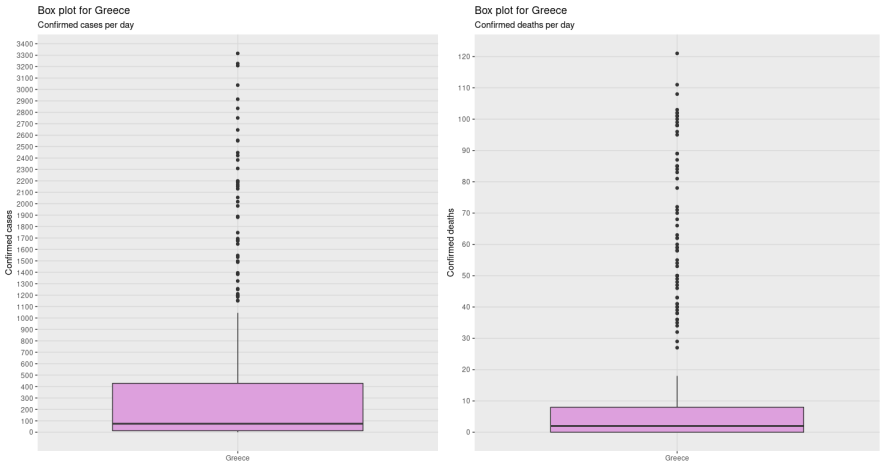

Figure 2.17: Boxplot Greece deaths and confirmed cases