



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Προγραμματιστικά Εργαλεία και Τεχνολογίες για Επιστήμη Δεδομένων

Exploratory Data Analysis using R

Ορφανουδάκης Φίλιππος Σκόβελεφ

phil.orfa@gmail.com

AM: 03400107

MSc , ΕΔΕΜΜ ΔΠΜΣ

Αθήνα, Ιανουάριος 2021

# 1 Εισαγωγή

Στο πλαίσιο της εργασίας θα παρουσιάσουμε με συνοπτικό αλλά ουσιαστικό τρόπο τα δεδομένα της εκφώνησης αλλά και ορισμένα πορίσματα που εξήγαμε μετά από επεξεργασία και μελέτη των 2 αρχείων μας. Ζητούμενο της εργασίας δεν αποτελεί η αναλυτική επεξήγηση του κώδικα και για αυτό τον λόγο δεν θα επεκταθούμε ιδιαίτερα σε αυτόν. Για λόγους χωρικής οικονομίας θα παραθέτουμε μια φορά τα αντίστοιχα κομμάτια που επαναλαμβάνονται.

Η συλλογιστική πορεία που ακολουθεί η εργασία και θα βοηθήσει τον αναγνώστη, αποτελείται από την αρχική κατανόηση των δεδομένων, από την παγκόσμια αναπαράσταση των δεδομένων και τέλος δραστηριοποιούμαστε πιο εντατικά στην Ευρώπη για την οποία μελετάμε σε μεγαλύτερο βάθος ορισμένες χώρες.

Τέλος να αναφέρουμε ότι δεν έχουμε στόχο να παραθέσουμε προσωπικές ιατρικές ή πολιτικές απόψεις.

## 2 Βασική Προ επεξεργασία

Αρχικά ξεκινήσαμε με το να διαβάσουμε τα δυο δοθέντα datasets και προχωρήσαμε στα πρώτα βήματα προ επεξεργασίας που όριζε η εργασία και αναφέρουμε αρκετά συνοπτικά κάποια επεξηγηματικά σχόλια:

### 1. Remove columns with names Province, State, Lat and Long.

```
1 library(data.table)
2 library(ggplot2)
3 library(dplyr)
4
5 rm(list = ls())
6
7 covid1 <- read.csv(file = 'time_series_covid19_confirmed_global.csv')
8 covid2 <- read.csv(file = 'time_series_covid19_deaths_global.csv')
9
10 covid_confirmed <- as.data.table(covid1)
11 covid_global <- as.data.table(covid2)
12
13
14 covid_confirmed[, c("Province.State", "Lat", "Long") := NULL]
15 covid_global[, c("Province.State", "Lat", "Long") := NULL]
16
17 names(covid_confirmed) <- gsub("X", "", names(covid_confirmed), fixed = TRUE)
18 names(covid_global) <- gsub("X", "", names(covid_global), fixed = TRUE)
```

### Σχόλια

Διαβάζουμε τα dataset και τα μετατρέπουμε σε datatables, στη συνέχεια αφαιρούμε τις στήλες που δεν χρειαζόμαστε και τέλος παρατηρούμε ότι κάθε στήλη αναφέρεται σε ημερομηνία αλλά μπροστά έχει το γράμμα "X", το οποίο και αφαιρούμε καθώς σε μελλοντική επεξεργασία μπορεί να φανεί εμπόδιο

### 2. Convert data from wide to long format.

```
1 colsc <- colnames(covid_confirmed)
2 colsg <- colnames(covid_global)
3
4 covid_confirmed.m = melt(covid_confirmed, id.vars = colnames(covid_confirmed[,1]),
5                           measure.vars = colnames(covid_confirmed[,2:length(colsc)]))
6
7 covid_global.m = melt(covid_global, id.vars = colnames(covid_global[,1]),
8                       measure.vars = colnames(covid_global[,2:length(colsg)]))
```

### Σχόλια

Με την μετατροπή αυτή πλέον κάθε στήλη-ημερομηνία μετατρέπεται σε τιμή μιας στήλης

### 3. Rename variable Country.Region to Country.

```
1 setnames(covid_confirmed.m, "Country.Region", "Country")
2 setnames(covid_global.m, "Country.Region", "Country")
```

#### Σχόλια

Απλή Μετονομασία

### 4. Name the variable with the cumulative confirmed cases as confirmed and the variable with the cumulative number of deaths as deaths.

```
1 setnames(covid_confirmed.m, "value", "confirmed")
2 setnames(covid_global.m, "value", "deaths")
```

#### Σχόλια

Απλή Μετονομασία

### 5. Convert the variable date from character to a date object

```
1 setnames(covid_confirmed.m, "variable", "date")
2 setnames(covid_global.m, "variable", "date")
3
4 covid_confirmed.m$date <- as.Date(covid_confirmed.m$date, format = "%m.%d.%y")
5 covid_global.m$date <- as.Date(covid_global.m$date, format = "%m.%d.%y")
```

#### Σχόλια

Μετατρέπουμε τις ημερομηνίες στην εξής μορφή : YY/MM/DD

### 6. Group by country and date.

```
1 covid_confirmed.m <- covid_confirmed.m %>% group_by(Country, date) %>% summarise(confirmed =
  sum(confirmed))
2 covid_global.m <- covid_global.m %>% group_by(Country, date) %>% summarise(deaths = sum(
  deaths))
```

#### Σχόλια

Εφαρμογή του group by και του summarize έτσι ώστε να μην έχουμε duplicates στην στήλη Country και date, και τα δεδομένα μας να είναι ομαδοποιημένα ανά περιοχή και ημερομηνία.

### 7. Merge the two datasets into one.

```
1 covid <- merge(covid_confirmed.m, covid_global.m, by=c("Country", "date"))
2 covid <- as.data.table(covid)
```

#### Σχόλια

Φτιάχνουμε ένα data.table ως συνένωση των 2 προηγούμενων

### 8. Calculate counts (confirmed and deaths) for the whole world.

```
1 world <- covid[covid$date==max(covid$date),]
2 world[, date:=NULL]
```

## Σχόλια

Για κάθε χώρα έχουμε τα συνολικά επιβεβαιωμένα κρούσματα και τους συνολικούς θανάτους, οι οποίοι αντιστοιχούσαν στην "μέγιστη" ημερομηνία των cumulative δεδομένων.

## 9. Sort (again) by country and date.

## Σχόλια

Έχει πραγματοποιηθεί ήδη

## 10. Create two extra variables: confirmed.inc and deaths.inc

```
1 covid <- covid%>%group_by(Country)%>%mutate(confirmed.inc = confirmed - lag(confirmed))
2 covid <- covid%>%group_by(Country)%>%mutate(deaths.inc = deaths - lag(deaths))
3
4 covid[covid$date==min(covid$date),5] <- covid[covid$date==min(covid$date),3]
5 covid[covid$date==min(covid$date),6] <- covid[covid$date==min(covid$date),4]
```

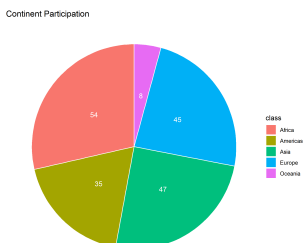
## Σχόλια

Δημιουργούμε μια νέα μεταβλητή με τα καθημερινά κρούσματα και θανάτους και αντικαθιστούμε τις NA τιμές στο data.table.

# 3 Βασικά στοιχεία των Δεδομένων

Πριν ξεκινήσουμε την εξερεύνηση και την αναπαράσταση των δεδομένων μας είναι χρήσιμο να κατανοήσουμε την φύση τους. Αρχικά πραγματευόμαστε δεδομένα που αφορούν επιβεβαιωμένα κρούσματα αλλά και θανάτους λόγω του COVID-19. Οι πληροφορίες που έχουμε στα χέρια μας και είμαστε ελεύθεροι να επεξεργαστούμε αφορούν 4 μεταβλητές:

- Καθημερινά επιβεβαιωμένα κρούσματα ανά χώρα
- Καθημερινοί θάνατοι ανά χώρα
- Συσσωρευτικά επιβεβαιωμένα κρούσματα ανά χώρα
- Συσσωρευτικοί θάνατοι ανά χώρα



**Figure 1:** Number of Countries per Continent

Πιο συγκεκριμένα έχουμε 189 χώρες που ανήκουν στις 5 ηπείρους και έχουμε και 2 πλοία, το Diamond Princess και το MS Zaandam τα οποία εντοπίστηκαν στο dataset. Για την γεωχωρική επεξεργασία αποφασίσαμε να μην ασχοληθούμε με τις παραπάνω περιπτώσεις.

Παραθέτουμε τα διαγράμματα στα οποία απεικονίζουμε τις πρώτες 5 χώρες με βάση τα επιβεβαιωμένα κρούσματα σε κάθε ήπειρο και βλέπουμε ότι μόνο στην Ευρώπη ο ιός επεκτάθηκε με τον ίδιο ρυθμό, φυσικά αυτό οφείλεται στην έκταση αλλά και τον πληθυσμό των χωρών. Ο αριθμός των θανάτων είναι ανάλογος με τα κρούσματα.



**Figure 2: Top 5 countries per Continent**

```

1 num_countries<-nrow(world)
2 num_countries
3 world$continent <- countrycode(sourcevar = world$Country, origin = "country.name",
  destination = "continent")
4 world$continent[world$Country=='Diamond Princess']='Ship'
5 world$continent[world$Country=='MS Zaandam']='Ship'
6 world$continent[world$Country=='Kosovo']='Europe'
7 Europe<-world[continent=="Europe",.N]
8 Asia<-world[continent=="Asia",.N]
9 Africa<-world[continent=="Africa",.N]
10 Oceania<-world[continent=="Oceania",.N]
11 Americas<-world[continent=="Americas",.N]
12 countr <- data.table(
13   group=c("Europe","Asia","Africa","Oceania","Americas"),
14   value=c(Europe,Asia,Africa,Oceania,Americas)
15 )
16 countr <- countr %>%
17   arrange(desc(group)) %>%
18   mutate(lab = cumsum(value) - 0.5*value)
19 ggplot(countr, aes(x="", y=value, fill=group)) +
20   geom_bar(stat="identity", width=1,color="white") +
21   labs(fill="class",
22     x=NULL,
23     y=NULL,
24     title="Continent Participation")+
25   geom_text(aes(y=lab,label = value),color="white")+
26   coord_polar("y", start=0)+
27   theme_void()
28 world_sorted_deaths<-world[order(continent,-deaths)]
29 top5cases_per_continent<-world%>%
30   group_by(continent) %>%
31   arrange(desc(confirmed)) %>%
32   slice(1:5)
33 topEuropecases<-top5cases_per_continent[top5cases_per_continent$continent=="Europe",]
34 topAfricacases<-top5cases_per_continent[top5cases_per_continent$continent=="Africa",]
35 topAsiacases<-top5cases_per_continent[top5cases_per_continent$continent=="Asia",]
36 topAmericascases<-top5cases_per_continent[top5cases_per_continent$continent=="Americas",]
37 topOceaniacases<-top5cases_per_continent[top5cases_per_continent$continent=="Oceania",]
38 meltAfrica<-melt.data.table(as.data.table(topAfricacases), id.vars = c("Country", "
  continent"), measure.vars = c("deaths", "confirmed"))
39 meltEurope<-melt.data.table(as.data.table(topEuropecases), id.vars = c("Country", "
  continent"), measure.vars = c("deaths", "confirmed"))
40 meltOceania<-melt.data.table(as.data.table(topOceaniacases), id.vars = c("Country", "
  continent"), measure.vars = c("deaths", "confirmed"))
41 meltAmericas<-melt.data.table(as.data.table(topAmericascases), id.vars = c("Country", "
  continent"), measure.vars = c("deaths", "confirmed"))
42 meltAsia<-melt.data.table(as.data.table(topAsiacases), id.vars = c("Country", "continent"),
  measure.vars = c("deaths", "confirmed"))
43 plotAfrica<-ggplot(meltAfrica, aes(fill=meltAfrica$variable, y=meltAfrica$value, x=
  meltAfrica$Country)) +
44   geom_bar(position="stack", stat="identity",alpha=.6, width=.4) +
45   theme_bw() +
46   xlab("")+
47   labs(title = "Africa",
48     y = "", x = "", fill = "")+
49   coord_flip()

```

## 4 Παγκόσμια εξέλιξη στο χρόνο

Έχοντας πλέον μια μικρή εικόνα των δεδομένων θα εξετάσουμε την χρονική εξέλιξη των κρουσμάτων αλλά και θανάτων σε παγκόσμιο επίπεδο. Στο σημείο αυτό να αναφέρουμε ότι τα δεδομένα καλύπτουν από 2020-01-22 έως και 2021-01-04

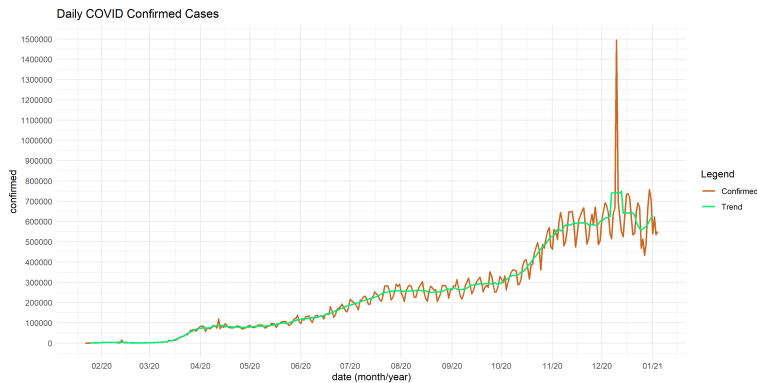


Figure 3: Daily Confirmed Cases

χαρακτηρίζεται από αυξητική τάση τόσο στα επιβεβαιωμένα κρούσματα όσο και στους θανάτους.

Τέλος θα σχολιάσουμε την κορυφή που εμφανίζει η χρονοσειρά των επιβεβαιωμένων κρουσμάτων. Πιο συγκεκριμένα στις 10 Δεκεμβρίου η Τουρκία επιβεβαίωσε 823225 κρούσματα, δηλαδή το σύνολο όλων των κρουσμάτων της μέχρι εκείνη την χρονική περίοδο. Το συγκεκριμένο ζήτημα αναφέρεται και στο [github](#) των dataset που χρησιμοποιούμε.

Σημείωση: Για τον υπολογισμό της τάσης χρησιμοποιήσαμε το moving average για διάστημα 7 ημερών.

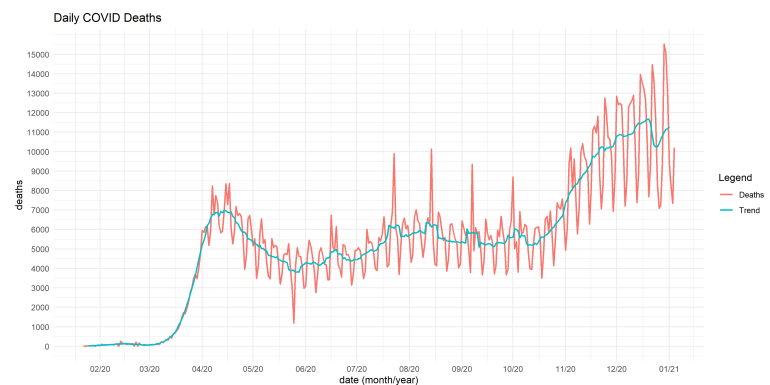


Figure 4: Daily Deaths

```
1 agg <- covid %>% group_by(date) %>% summarize(confirmed.inc=sum(confirmed.inc), deaths.inc=
  sum(deaths.inc)) %>% mutate(days = date - first(date) + 1)
2 agg <- agg %>% mutate(rolling_conf = rollmean(confirmed.inc, k = 7, fill = NA))
3 agg <- agg %>% mutate(rolling_death = rollmean(deaths.inc, k = 7, fill = NA))
4
5 agg<-as.data.table(agg)
6
7 ggplot(mapping = aes()) +
8   geom_line(data = agg, mapping = aes(x=date, y=deaths.inc,color="Deaths"),size=0.8) +
9   geom_line(data = agg, mapping = aes(x=date, y=rolling_death,color="Trend"),size=0.8)+
10  geom_point() +
11  labs(x = "date (month/year)",
12       y = "deaths",
13       color = "Legend",
14       title = "Daily COVID Deaths") +
15  scale_x_date(breaks = date_breaks("months"), labels = date_format("%m/%y")) +
16  scale_y_continuous(breaks = seq(0,max(agg$deaths.inc), by=1000)) +
```

```

17 theme_minimal()
18
19 ggplot(mapping = aes()) +
20   geom_line(data = agg, mapping = aes(x=date, y=confirmed.ind,color="Confirmed"),size=0.8)
21   +
22   geom_line(data = agg, mapping = aes(x=date, y=rolling_conf,color="Trend"),size=0.8)+
23   labs(x = "date (month/year)",
24        y = "confirmed",
25        color="Legend",
26        title = "Daily COVID Confirmed Cases") +
27   scale_x_date(breaks = date_breaks("months"), labels = date_format("%m/%y")) +
28   scale_y_continuous(breaks = seq(0,max(agg$confirmed.ind)+10000, by=100000)) +
29   scale_color_manual(values=c("chocolate3", "springgreen2"))+
30   theme_minimal()

```

## 5 Πανδημία στην Ευρώπη

Σαν κεντρικό θέμα αυτής της εργασίας αποφασίσαμε να μελετήσουμε την Ευρώπη για συγκεκριμένους λόγους. Αρχικά σαν χώρα βρισκόμαστε στην Ευρώπη και είναι ιδιαίτερα ενδιαφέρον να συγκρίνουμε την πορεία της Ελλάδας με άλλες ευρωπαϊκές χώρες. Στη συνέχεια, αναφέραμε στην ενότητα 3 ότι από όλες τις ηπείρους η Ευρώπη είχε ομοιόμορφη κατανομή στην εξάπλωση του ιού στις χώρες της, επομένως μπορούμε να εξετάσουμε ομοιότητες και διαφορές μεταξύ αυτών των χωρών. Τέλος ορισμένες ευρωπαϊκές χώρες δυστυχώς μονοπώλησαν την προσοχή καθ' όλη την διάρκεια της πανδημίας και θέλαμε να εξετάσουμε αυτές τις περιπτώσεις για να αποκτήσουμε μια καλύτερη άποψη. Αρχικά παραθέτουμε την κατάταξη των ευρωπαϊκών χωρών σχετικά με τα επιβεβαιωμένα κρούσματα αλλά και τους θανάτους.

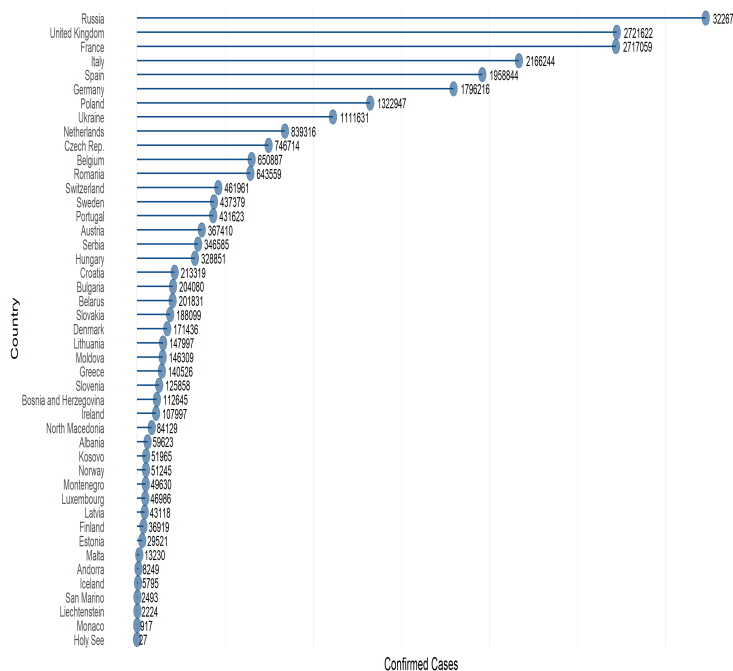


Figure 5: Confirmed Cases

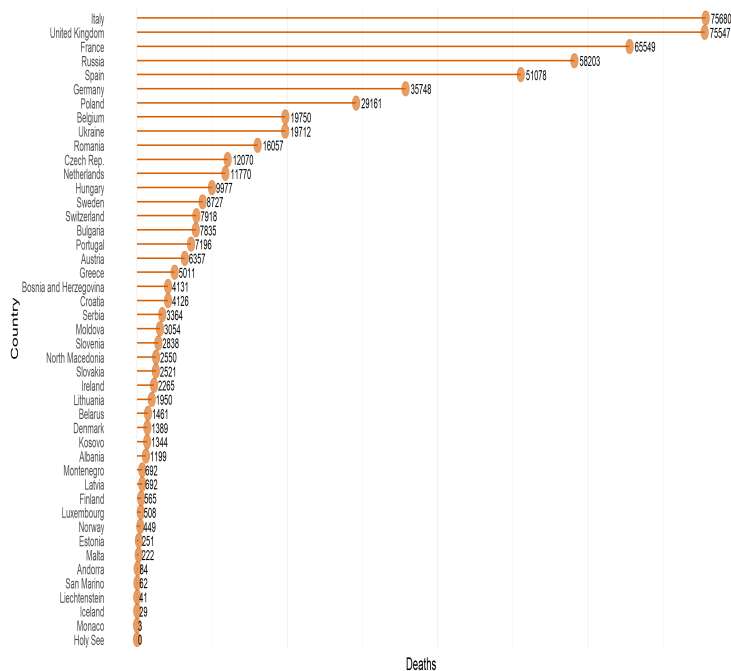


Figure 6: Deaths

Παρατηρούμε ότι στις πρώτες 5 θέσεις (με διαφορετική σειρά) βρίσκονται οι : **Ρωσία, Ισπανία, Ηνωμένο Βασίλειο, Ιταλία, Γαλλία**. Ενώ η Ελλάδα βρίσκεται στην 26η θέση για τα επιβεβαιωμένα κρούσματα και στη 19η θέση για τους θανάτους, γεγονός που δείχνει ένα επιβαρυνμένο υγειονομικό σύστημα. Από αυτές τις χώρες επιλέγουμε να κοιτάξουμε σε μεγαλύτερο βάθος την Ισπανία, Ιταλία, Ηνωμένο Βασίλειο και Ελλάδα. Θεωρούμε ότι αυτές οι χώρες μονοπώλησαν μέχρι τώρα το ενδιαφέρον μας.

Παραθέτουμε 2 χάρτες που μας δείχνουν τον αριθμό των κρουσμάτων και θανάτων διαιρεμένο με τον πληθυσμό της χώρας για να αποκτήσουμε μια εποπτική γνώση για την σοβαρότητα των χωρών που επιλέξαμε.



Confirmed Cases per Population

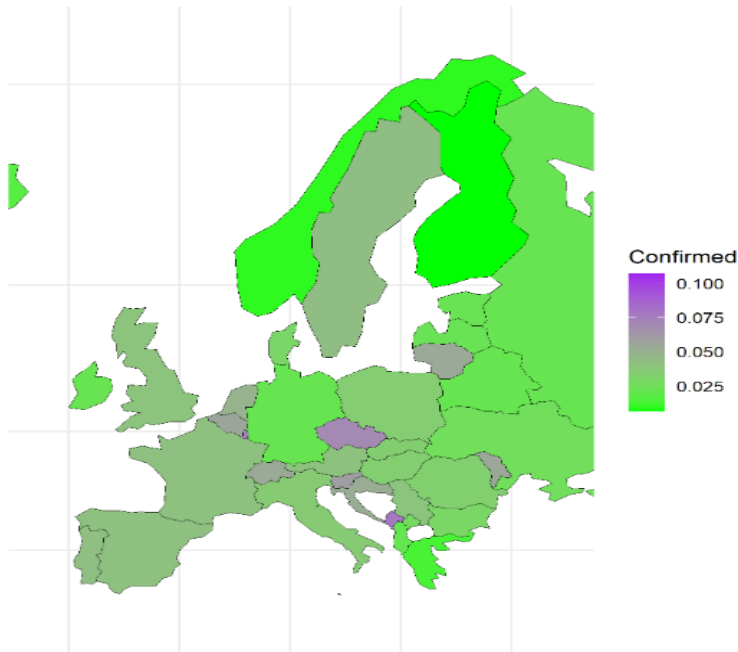


Figure 7: Confirmed Cases HeatMap

Deaths per Population

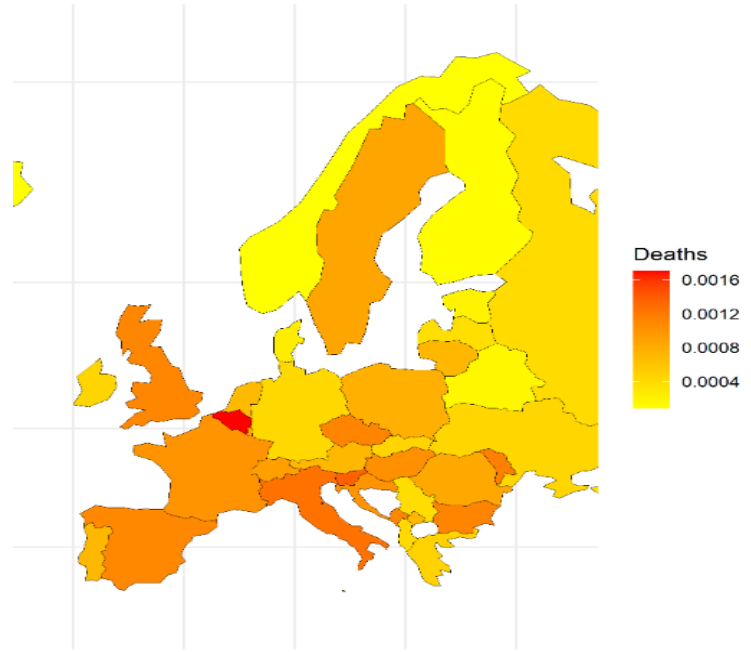


Figure 8: Deaths HeatMap

Παρατηρούμε ότι η Ισπανία και το Ηνωμένο Βασίλειο ανήκουν στη κατηγορία υψηλής σοβαρότητας, τόσο για τα επιβεβαιωμένα κρούσματα, όσο και για τους θανάτους. Η Ιταλία παρουσιάζει μια κατάσταση ελαφρώς πιο ελεγχόμενη για τα κρούσματα αλλά αρκετά πιο σοβαρή για τους θανάτους. Από την άλλη, η Ελλάδα φαίνεται πως δεν είχε το αντίστοιχο "πλήγμα" με τις χώρες που αναφέρθηκαν παραπάνω.

Σύμφωνα με τους χάρτες, η χώρα που βρίσκεται στην πιο κρίσιμη κατάσταση σχετικά με τους θανάτους είναι το Βέλγιο ενώ η Τσεχία είναι ιδιαίτερα επιβαρυμένη σχετικά με τα κρούσματα. Πιο συγκεκριμένα :

- Βέλγιο : 650887 επιβεβαιωμένα κρούσματα και 19750 θάνατοι
- Τσεχία : 746714 επιβεβαιωμένα κρούσματα και 12070 θάνατοι

```

1 worldMap <- getMap()
2
3
4 data(pop)
5 population<-pop[,c("name", "2020")]
6 setnames(population,"2020","twenty")
7 world$pop<-population$twenty[match(world$Country,population$name)]
8
9 world[Country=="Czechia"]$Country="Czech Rep."
10 world[Country=="Russia"]$pop=145934.460
11 world[Country=="Andorra"]$pop=77.006
12 world[Country=="Czech Rep."]$pop=10708.982
13 world[Country=="Kosovo"]$pop=1873.160
14 world[Country=="Moldova"]$pop=2640.438
15
16
17 world <- transform(world, c_ratio =confirmed /(1000*pop))
18 world <- transform(world, d_ratio =deaths /(1000*pop))
19 eu<-world[continent=="Europe"]$Country
20
21 indEU <- which(worldMap$NAME%in%eu)
22 europeCoords <- lapply(indEU, function(i){
23   df <- data.frame(worldMap@polygons[[i]]@Polygons[[1]]@coords)
24   df$region =as.character(worldMap$NAME[i])
25   colnames(df) <- list("long", "lat", "region")

```

```

26   return(df)
27 })
28
29 world[Country=="Czech Rep."]
30 world[Country=="Belgium"]
31
32
33 europeCoords <- do.call("rbind", europeCoords)
34 europeCoords$Deaths <- world$d_ratio[match(europeCoords$region,world$Country)]
35 europeCoords$Cases<-world$c_ratio[match(europeCoords$region,world$Country)]
36
37
38
39
40 P <- ggplot() + geom_polygon(data = europeCoords, aes(x = long, y = lat, group = region,
   fill = Deaths),
41                               colour = "black", size = 0.1) +
42   coord_map(xlim = c(-13, 35), ylim = c(32, 71))+
43   theme_minimal()+
44   scale_fill_gradient(name = "Deaths", low = "#FFFF00FF", high = "#FF0000FF", na.value = "
   grey50")+
45   theme(#panel.grid.minor = element_line(colour = NA), panel.grid.minor = element_line(
   colour = NA),
46         #panel.background = element_rect(fill = NA, colour = NA),
47         axis.text.x = element_blank(),
48         axis.text.y = element_blank(), axis.ticks.x = element_blank(),
49         axis.ticks.y = element_blank(), axis.title = element_blank(),
50         #rect = element_blank(),
51         plot.margin = unit(0 * c(-1.5, -1.5, -1.5, -1.5), "lines"))+
52   labs(title = "Deaths per Population ")
53
54 P
55 ggsave('R_images/death_ratio.png', width = width, height = height, units="cm")
56
57 P1<- ggplot() + geom_polygon(data = europeCoords, aes(x = long, y = lat, group = region,
   fill = Cases),
58                               colour = "black", size = 0.1) +
59   coord_map(xlim = c(-13, 35), ylim = c(32, 71))+
60   theme_minimal()+
61   scale_fill_gradient(name = "Confirmed", low = "green", high = "purple", na.value = "white
   ")+
62   theme(#panel.grid.minor = element_line(colour = NA), panel.grid.minor = element_line(
   colour = NA),
63         #panel.background = element_rect(fill = NA, colour = NA),
64         axis.text.x = element_blank(),
65         axis.text.y = element_blank(), axis.ticks.x = element_blank(),
66         axis.ticks.y = element_blank(), axis.title = element_blank(),
67         #rect = element_blank(),
68         plot.margin = unit(0 * c(-1.5, -1.5, -1.5, -1.5), "lines"))+
69   labs(title = "Confirmed Cases per Population")
70 P1
71 ggsave('R_images/cases_ratio.png', width = width, height = height, units="cm")
72
73
74 Europe<-world[continent=="Europe"]
75 Europe<-Europe[order(confirmed)]
76
77 Europe_plot_cases<-Europe %>%
78   arrange(confirmed) %>%
79   mutate(state = factor(Country, unique(Country))) %>%
80   ggplot() + aes(x=Country, y=confirmed) +
81   geom_segment( aes(x=state, xend=state, y=0, yend=confirmed), color="dodgerblue4") +

```

```

82 geom_point( color="dodgerblue4", size=4, alpha=0.6) +
83 geom_text(aes(label = confirmed), hjust = -0.25,size=3) +
84 theme_light() +
85 coord_flip() +
86 theme(
87   panel.grid.major.y = element_blank(),
88   panel.border = element_blank(),
89   axis.ticks.y = element_blank(),
90   axis.text.x=element_blank(),
91   axis.ticks.x=element_blank()
92 )+
93 labs(y = "Confirmed Cases")
94
95 Europe_plot_cases
96 ggsave('R_images/Europe_cases_rank.png', width = width, height = height, units="cm")
97
98 Europe_plot_Deaths<-Europe %>%
99   arrange(deaths) %>%
100   mutate(state = factor(Country, unique(Country))) %>%
101   ggplot() + aes(x=Country, y=deaths) +
102   geom_segment( aes(x=state, xend=state, y=0, yend=deaths), color="#D55E00") +
103   geom_point( color="#D55E00", size=4, alpha=0.6) +
104   geom_text(aes(label = deaths), hjust = -0.25,size=3) +
105   theme_light() +
106   coord_flip() +
107   theme(
108     panel.grid.major.y = element_blank(),
109     panel.border = element_blank(),
110     axis.ticks.y = element_blank(),
111     axis.text.x=element_blank(),
112     axis.ticks.x=element_blank()
113 )+
114   labs(y = "Deaths")
115
116 Europe_plot_Deaths
117 ggsave('R_images/Europe_deaths_rank.png', width = width, height = height, units="cm")

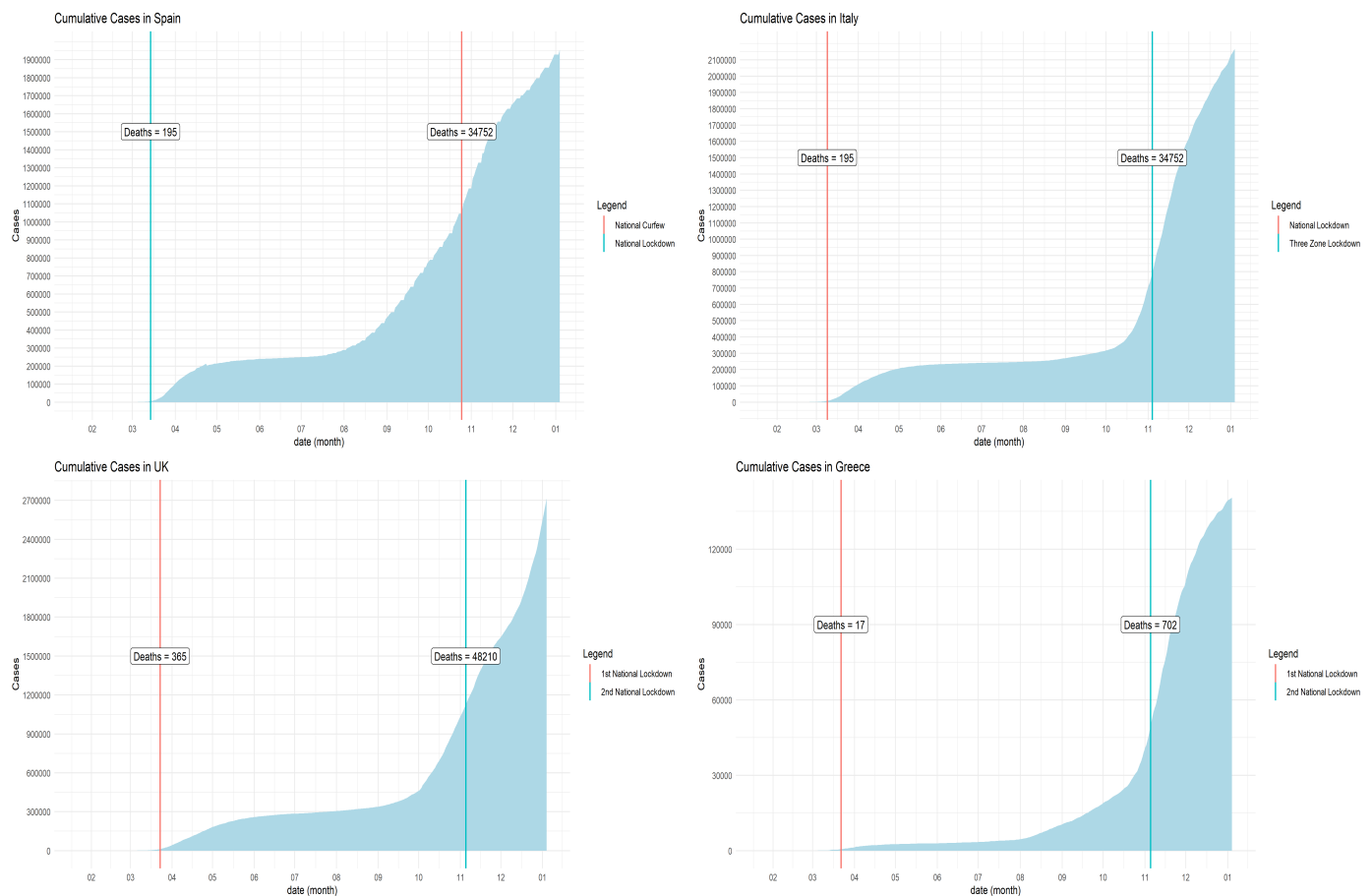
```

## 6 Μελέτη Ισπανίας, Ιταλίας, Ηνωμένου Βασιλείου και Ελλάδας

Για κάθε μια από αυτές τις χώρες θα εξετάσουμε 3 συγκεκριμένα ζητούμενα :

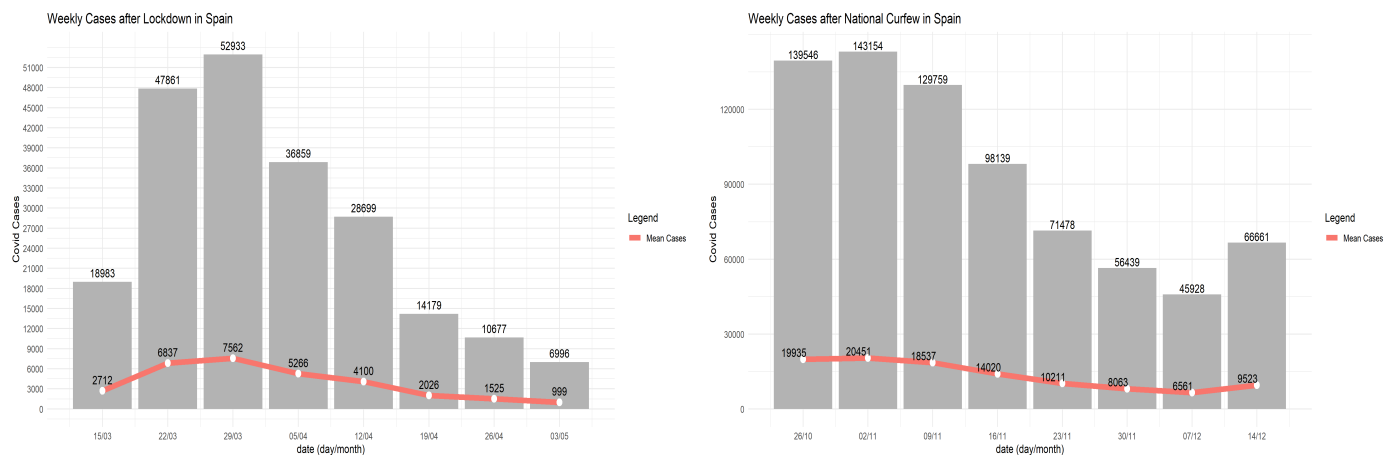
- Συσσωρευτικά επιβεβαιωμένα κρούσματα με αναφορά σε 2 κρίσιμες ημερομηνίες (πχ. ανακοίνωση lockdown, απαγόρευση κυκλοφορίας).
- Πορεία κρουσμάτων για τους επόμενους 2 μήνες έπειτα από κάθε μια από τις 2 κρίσιμες ημερομηνίες
- Επιβεβαιωμένα Κρούσματα για κάθε μέρα της εβδομάδας

Ξεκινώντας με το πρώτο ζητούμενο παραθέτουμε τα διαγράμματα. Παρατηρούμε τεράστια ομοιότητα τόσο στη πορεία των κρουσμάτων όσο και στη στιγμή που τέθηκαν αυστηρότερα μέτρα. Μπορούμε να αναγνωρίσουμε ότι τα πρώτα μέτρα είχαν φανερά αποτελέσματα και κατάφεραν να κάνουν την καμπύλη πιο επίπεδη. Δυστυχώς δεν μπορούμε να πούμε το ίδιο για τα δεύτερα μέτρα, ειδικά για το Ηνωμένο Βασίλειο που η κλίση της καμπύλης φαίνεται να αυξάνεται.



**Figure 9:** (a)Spain (b)Italy (c)UK (d)Greece

Στη συνέχεια παραθέτουμε τα διαγράμματα που αφορούν την μετέπειτα πορεία των κρουσμάτων για κάθε χώρα ύστερα από τις 2 αυτές κρίσιμες ημερομηνίες.



**Figure 10:** Spain

Τα διαγράμματα αυτά μας δίνουν μια ξεκάθαρη εικόνα για την πορεία κάθε χώρας. Η Ισπανία, η Ιταλία και η Ελλάδα ανταποκρίθηκαν στα μέτρα που επιβλήθηκαν και έχουμε πτωτική πορεία. Παρατηρούμε για αυτές τις χώρες ότι τα μέτρα που εφαρμόζονται έχουν σαν αποτέλεσμα ύστερα από 3 εβδομάδες να μειώνουν τα καθημερινά κρούσματα, ενώ τις πρώτες 3 εβδομάδες αυτά συνεχίζουν να αυξάνονται. Από την άλλη πλευρά στο Ηνωμένο Βασίλειο αποτυπώνεται μια απαισιόδοξη κατάσταση. Βλέπουμε μια σχετική σταθεροποίηση ή ακόμα και πτώση στα πρώτα μέτρα, αλλά στα δεύτερα μέτρα τα αποτελέσματα φαίνονται υποσχόμενα τον πρώτο μήνα και ύστερα τον Δεκέμβρη έχουμε ανοδική πορεία.

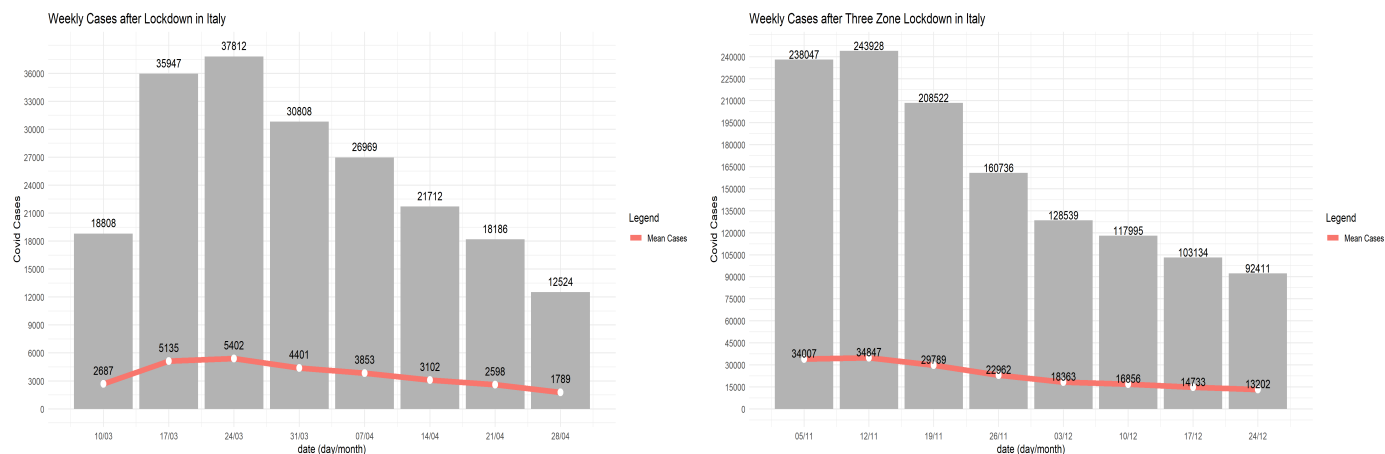


Figure 11: Italy

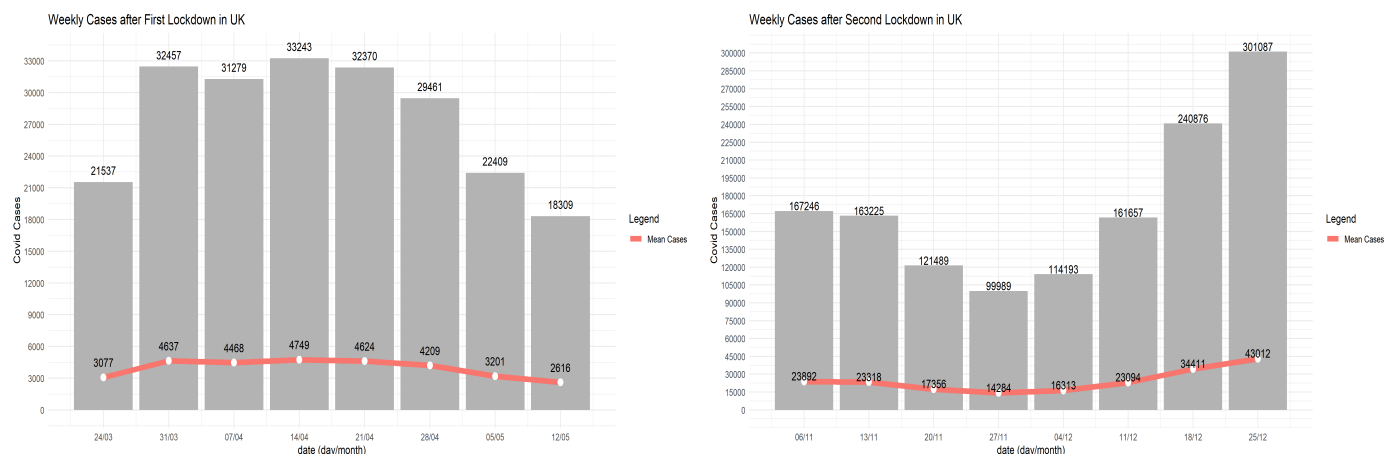


Figure 12: UK

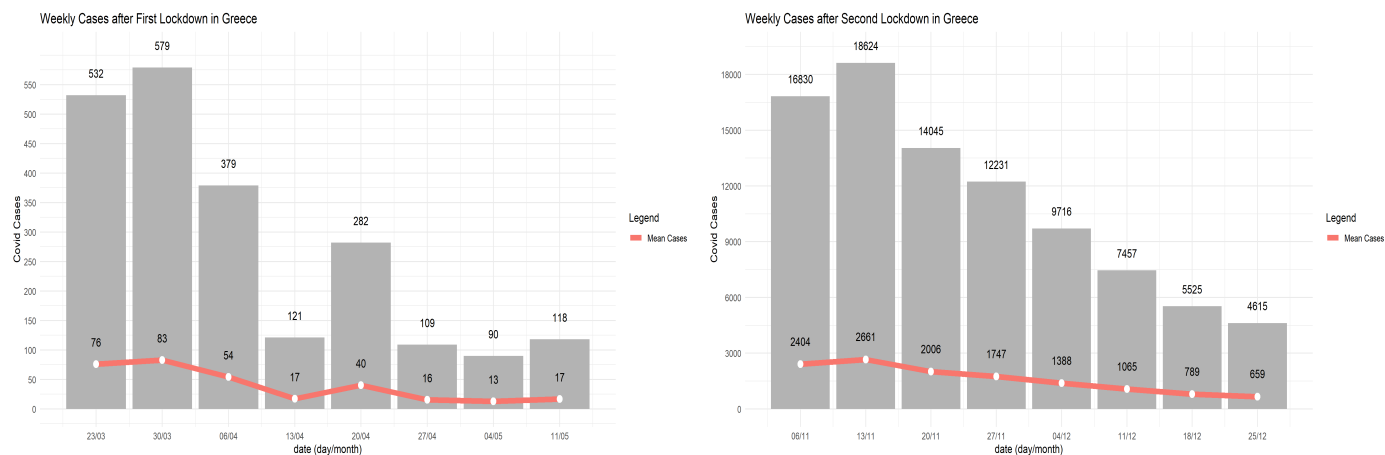
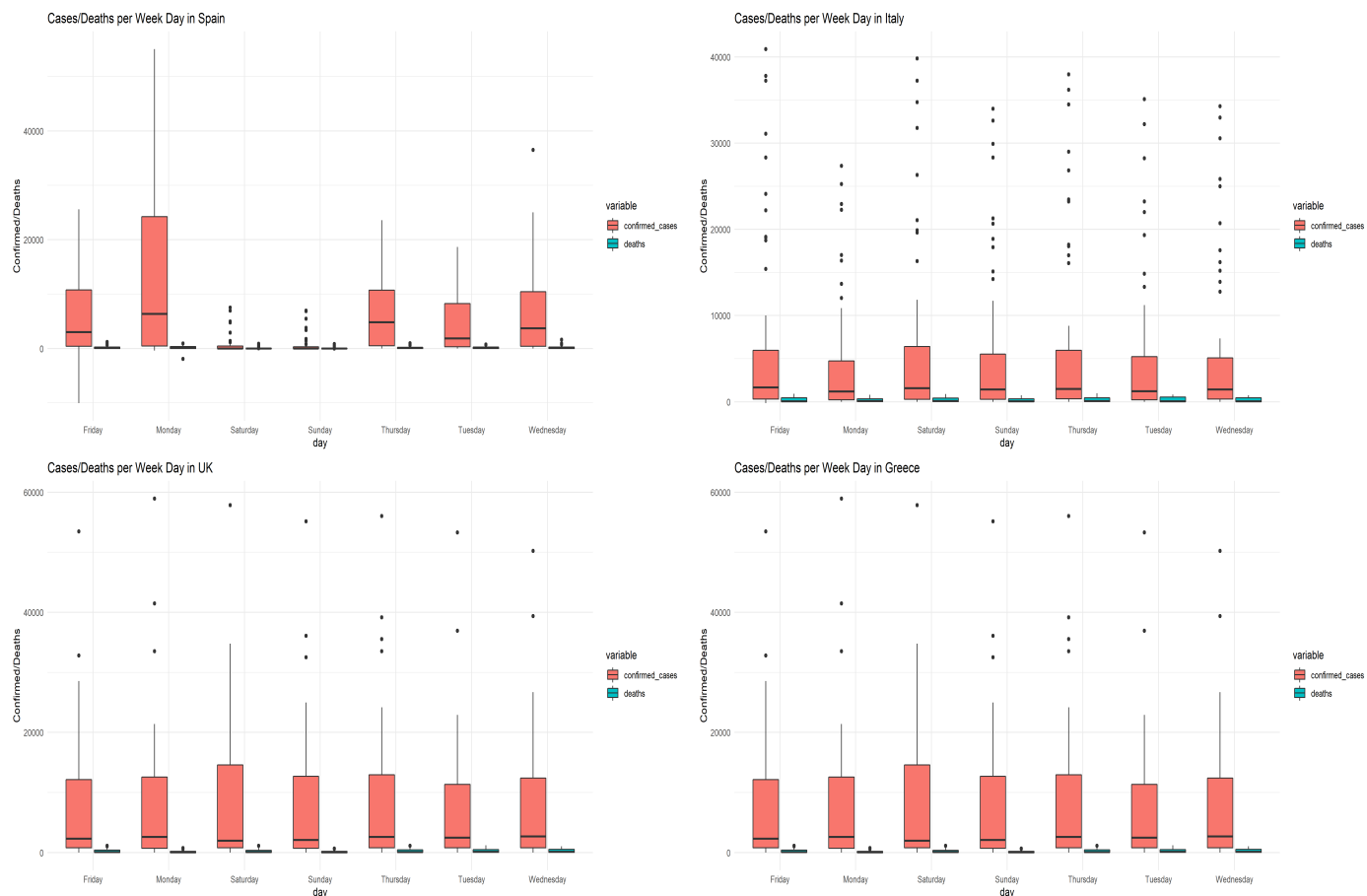


Figure 13: Greece

Τέλος, γίνεται συχνά αναφορά στη χώρα μας για τον σχετικά μικρό αριθμό ελέγχων που πραγματοποιούνται το Σάββατο και την Κυριακή και θελήσαμε να εκτιμήσουμε αυτόν τον αριθμό μέσα από τα κρούσματα που σημειώνονται για κάθε μια από τις μέρες της εβδομάδας. Παραθέτουμε τα αντίστοιχα διαγράμματα.

Αρκετά ενδιαφέρον είναι το γεγονός ότι η Ισπανία στη πλειοψηφία των περιπτώσεων δεν αναφέρει καθόλου κρούσματα και θανάτους το Σ/Κ, ενώ την Δευτέρα σημειώνεται μεγάλη άνοδος, πιθανότατα η μεγάλη άνοδος να συνεπάγεται και περισσότερα τεστ. Οι υπόλοιπες χώρες παρουσιάζουν αρκετά ομοιόμορφη εικόνα για τις ημέρες τις εβδομάδας, χωρίς να εντοπίζουμε κάποιο φαινόμενο για να επισημάνουμε.



**Figure 14:** (a)Spain (b)Italy (c)UK (d)Greece

```

1 Spain<-covid[covid$Country=="Spain",] %>% group_by(date) %>% summarize(confirmed.ind=sum(
  confirmed.ind), deaths.inc=sum(deaths.inc)) %>% mutate(days = date - first(date) + 1)
2
3 Spain<-as.data.table(Spain)
4 Spain[, ConfiCum := cumsum(confirmed.ind)]
5 Spain[, DeathsCum := cumsum(deaths.inc)]
6
7 S_lockdown1 = mdy("3/14/20")
8 S_lockdown2 = mdy("10/25/20")
9
10 S_Deaths_lockdown1<-Spain[date==S_lockdown1]$DeathsCum
11 S_Deaths_lockdown2<-Spain[date==S_lockdown2]$DeathsCum
12
13 Spain_Cases<-ggplot(mapping = aes()) +
14   geom_area(data = Spain, mapping = aes(x=date, y=ConfiCum),fill = "lightblue") +
15   geom_point() +
16   labs(x = "date (month)",
17        y = "Cases",
18        color = "Legend",
19        title = "Cumulative Cases in Spain") +
20   scale_x_date(breaks = date_breaks("months"), labels = date_format("%m")) +
21   geom_vline(size=0.8, mapping = aes(xintercept=as.numeric(S_lockdown1), color="National
    Lockdown"), show.legend = T) +
22   geom_vline(size=0.8, mapping = aes(xintercept=as.numeric(S_lockdown2), color="National
    Curfew"), show.legend = T)+
23   scale_y_continuous(breaks = seq(min(Spain$ConfiCum),max(Spain$ConfiCum), by=100000)) +
24   theme_minimal()+
25   annotate("label", x = S_lockdown1, y = 1500000, label = "Deaths = 195") +
26   annotate("label", x = S_lockdown2, y = 1500000, label = "Deaths = 34752")

```

```

28 Spain_Cases
29 ggsave('R_images/Spain_cases.png', width = width, height = height, units="cm")
30
31
32 Spainlock1<-as.data.table(Spain)[date>S_lockdown1 & date<S_lockdown1+60]
33 Spainlock2<-as.data.table(Spain)[date>S_lockdown2 & date<S_lockdown2+60]
34
35 Spainlock1<-Spainlock1 %>% group_by(week = floor_date(date, unit="week")) %>% summarise(
  mean_cases = mean(confirmed.ind),weekly_cases=sum(confirmed.ind))
36 Spainlock2<-Spainlock2 %>% group_by(week = floor_date(date-1, unit="week")) %>% summarise(
  mean_cases = mean(confirmed.ind),weekly_cases=sum(confirmed.ind))
37
38 #Spainlock1<-Spainlock1 %>% group_by(week = cut(date, "week",start.on.monday = FALSE)) %>%
  summarise(mean_cases = mean(confirmed.ind),weekly_cases=sum(confirmed.ind))
39
40 Spainlock1<-Spainlock1[1:8,]
41
42 #Spainlock2<-Spainlock2 %>% group_by(week = cut(date, "week",start.on.monday = TRUE)) %>%
  summarise(mean_cases = mean(confirmed.ind),weekly_cases=sum(confirmed.ind))
43
44 Spainlock2<-Spainlock2[1:8,]
45
46 Spain_lock1<-ggplot(mapping = aes()) +
47   geom_bar(data = Spainlock1, mapping = aes(x=as.Date(week), y=weekly_cases),'identity',
48     fill = "grey70") +
49   geom_line(data = Spainlock1, mapping = aes(x=as.Date(week), y=mean_cases,color="Mean
50     Cases"),size=2.5)+
51   geom_point( data=Spainlock1, aes(x= as.Date(week),y=mean_cases), size=3.0, colour="white"
52     ) +
53   geom_text(data=Spainlock1,aes(x=as.Date(week), y=weekly_cases+700,label=weekly_cases),
54     vjust=0)+
55   geom_text(data=Spainlock1,aes(x=as.Date(week), y=mean_cases+1000,label=round(mean_cases))
56     ,vjust=0)+
57   labs(x = "date (day/month)",
58     y = "Covid Cases",
59     color = "Legend",
60     title = "Weekly Cases after Lockdown in Spain") +
61   scale_x_date(breaks = seq(min(Spainlock1$week),max(Spainlock1$week),by=7), labels = date_
62     format("%d/%m")) +
63   scale_y_continuous(breaks = seq(0,max(Spainlock1$weekly_cases), by=3000)) +
64   theme_minimal()
65
66 Spain_lock1
67 ggsave('R_images/Spain_lock1.png', width = width, height = height, units="cm")
68
69
70 Spain_lock2<-ggplot(mapping = aes()) +
71   geom_bar(data = Spainlock2, mapping = aes(x=as.Date(week)+1, y=weekly_cases),'identity',
72     fill = "grey70") +
73   geom_line(data = Spainlock2, mapping = aes(x=as.Date(week)+1, y=mean_cases,color="Mean
74     Cases"),size=2.5)+
75   geom_point( data=Spainlock2, aes(x= as.Date(week)+1,y=mean_cases), size=3.0, colour="
76     white" ) +
77   geom_text(data=Spainlock2,aes(x=as.Date(week)+1, y=weekly_cases+700,label=weekly_cases),
78     vjust=0)+
79   geom_text(data=Spainlock2,aes(x=as.Date(week), y=mean_cases+1000,label=round(mean_cases))
80     ,vjust=0)+
81   labs(x = "date (day/month)",
82     y = "Covid Cases",
83     color = "Legend",
84     title = "Weekly Cases after National Curfew in Spain") +
85   scale_x_date(breaks = seq(min(Spainlock2$week)+1,max(Spainlock2$week)+1,by=7), labels =
86     date_format("%d/%m")) +

```

```

74   scale_y_continuous(breaks = seq(0,max(Spainlock2$weekly_cases), by=30000)) +
75   theme_minimal()
76
77 Spain_lock2
78
79 ggsave('R_images/Spain.png', width = width, height = height, units="cm")
80
81 Europe<-world[continent=="Europe"]
82 Europe<-Europe[order(confirmed)]
83
84 Spain$day <- weekdays(as.Date(Spain$date))
85
86 meltSpain<-melt.data.table(as.data.table(Spain), id.vars = c("date", "days","day","ConfCum",
87   "DeathsCum"))
88
89 S_boxplot<-ggplot(meltSpain, aes(x=day, y=value, fill=variable)) +
90   geom_boxplot()+
91   theme_minimal()+
92   labs(y = "Confirmed/Deaths")+
93   scale_fill_discrete(labels = c("confirmed_cases", "deaths"))+
94   labs(title = "Cases/Deaths per Week Day in Spain")
95
96 S_boxplot
97
98 ggsave('R_images/Spain_box.png', width = width, height = height, units="cm")

```

---

<sup>0</sup>Ολόκληρος ο κώδικας βρίσκεται στο προσωπικό μου [github](#), στα πλαίσια της εργασίας παραθέσαμε τα σημεία του κώδικα που θεωρήσαμε απαραίτητα, τα κομμάτια που αφαιρέθηκαν είναι η φόρτωση των βιβλιοθηκών και τα σημεία που επαναλαμβάνονται για διαφορετικές χώρες. Θεωρήσαμε ότι η αρίθμηση των σελίδων ξεκινάει μετά απο το εξώφυλλο