

Raport Projekt nr 2

Urszula Białończyk, Kacper Kurowski, Yevhenii Vinichenko

30 11 2020

Wprowadzenie

W ramach projektu nr 2 z TWD (Technik Wizualizacji Danych) pracowaliśmy nad wizualizacją danych związanych z pandemią wirusa SARS-CoV-2. Jako temat przewodni wybraliśmy “Śmierci i Hospitalizacje w krajach EU w związku z pandemią wirusa SARS-CoV-2”. Temat ten wybraliśmy gdyż mieliśmy pewne wstępne przekonania o tym jak wyglądają zależności między nimi i chcieliśmy sprawdzić, czy mają one pokrycie w dostępnych danych.

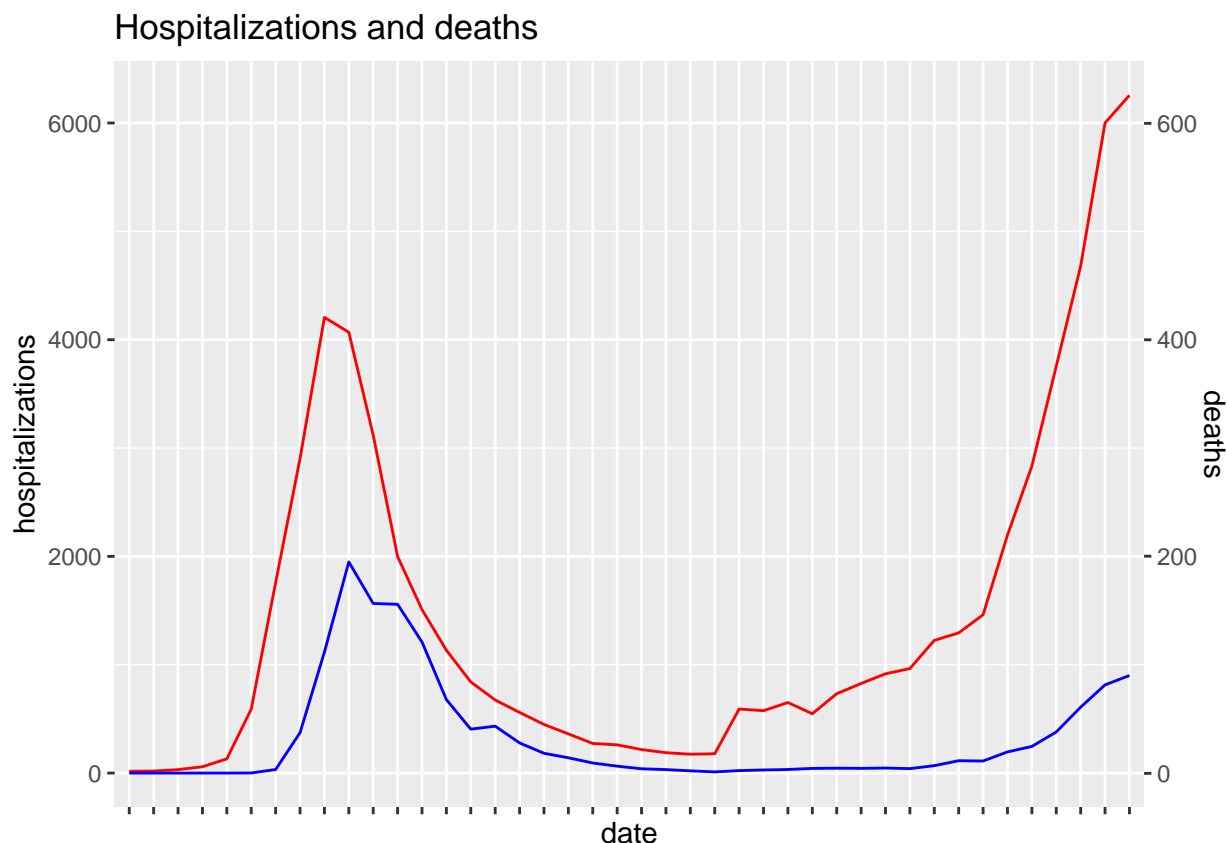
Porównanie liczby hospitalizacji i śmierci

Pierwsze wykresy dotyczą porównania liczby hospitalizacji i śmierci między krajami EU.

Porównanie liczby osób hospitalizowanych i śmierci danego dnia

Poniższy wykres przedstawia porównanie liczby osób hospitalizowanych i śmierci w danym tygodniu w krajach EU. Czerwonym kolorem zaznaczone hospitalizacje, niebieskim śmierci.

```
ggplot() +  
  geom_line(data = tmp, aes(x = date, y = hosp), group = 1, colour = 'red') +  
  geom_line(data = tmp, aes(x = date, y = deaths*10), group = 1, colour = 'blue') +  
  theme(axis.text.x = element_blank()) +  
  scale_y_continuous(  
    "hospitalizations",  
    sec.axis = sec_axis(~./10, name = "deaths")) +  
  labs(title = "Hospitalizations and deaths") +  
  scale_color_manual(values = c("red", "blue"))
```



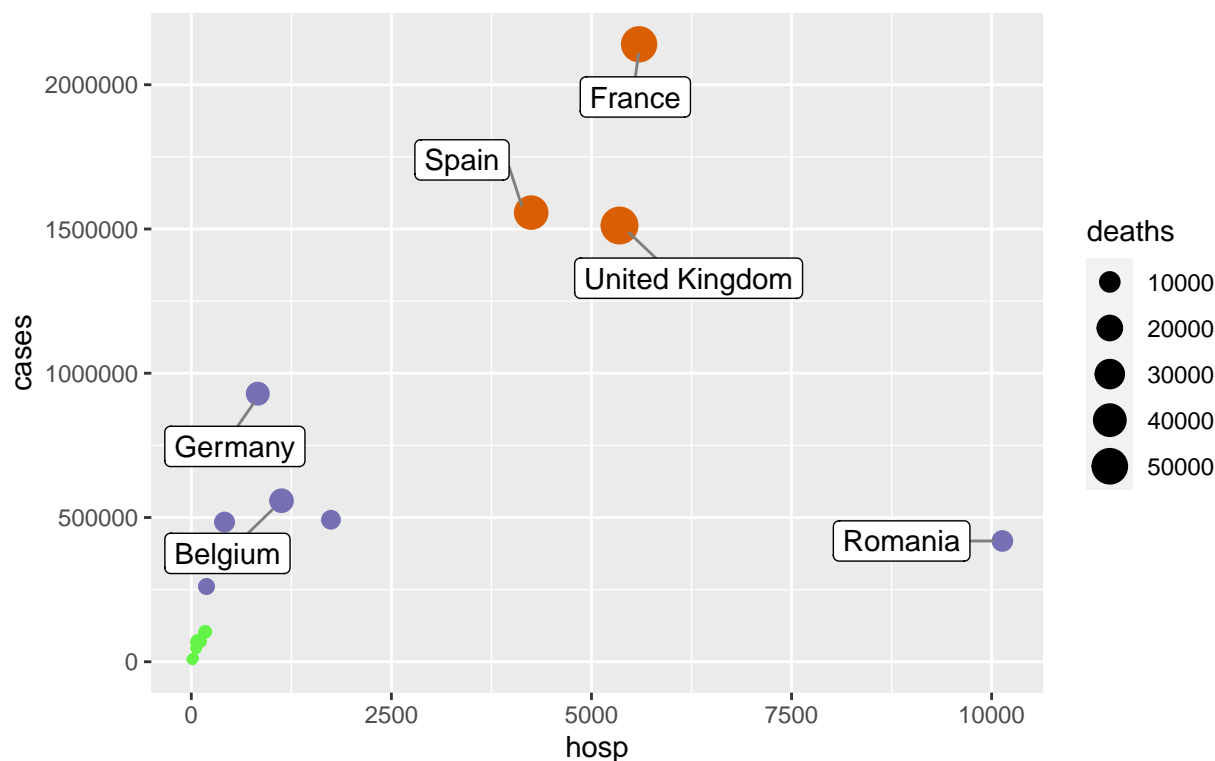
Podczas pierwszej fali widać jest, że prawie od razu po hospitalizacji dużej ilości osób zwiększyła się liczba śmierci, co mogło być spowodowane nieprzygotowaniem do wybuchu pandemii, zaś w czasie drugiej fali jest więcej hospitalizowanych i mniej śmierci, co świadczy o pewnym poziomie gotowości

Porównanie łącznej liczby osób ze stwierdzonym zakażeniem, hospitalizowanych, śmierci

Poniższy wykres przedstawia porównanie łącznej liczby osób ze stwierdzonym zakażeniem, hospitalizowanych i śmierci między krajami EU.

```
ggplot(data=tmp2, aes(x = hosp, y = cases, label = location)) +
  geom_point(aes(size = deaths), color = case_when(tmp2$cases > 10**6 ~ "#d95f02",
                                                    tmp2$cases < 2*10**5 ~ "#64f448",
                                                    TRUE ~ "#7570b3")) +
  geom_label_repel(data = subset(tmp2, deaths > 10000),
    aes(label = location),
    box.padding = 0.35,
    point.padding = 0.5,
    segment.color = 'grey50') +
  labs(title = "How do cases and hospitalizations\ninfluence deaths")
```

How do cases and hospitalizations influence deaths



Wykres sugeruje, że w krajach o dużej ilości przypadków i stosunkowo niewielkiej ilości hospitalizowanych jest dużo zgonów, w odróżnieniu od tych, gdzie hospitalizacji jest dużo nawet przy małej ilości zakażeń.

Porównanie śmierci z tego roku vs poprzednie lata

Porównanie liczby śmierci między krajami EU

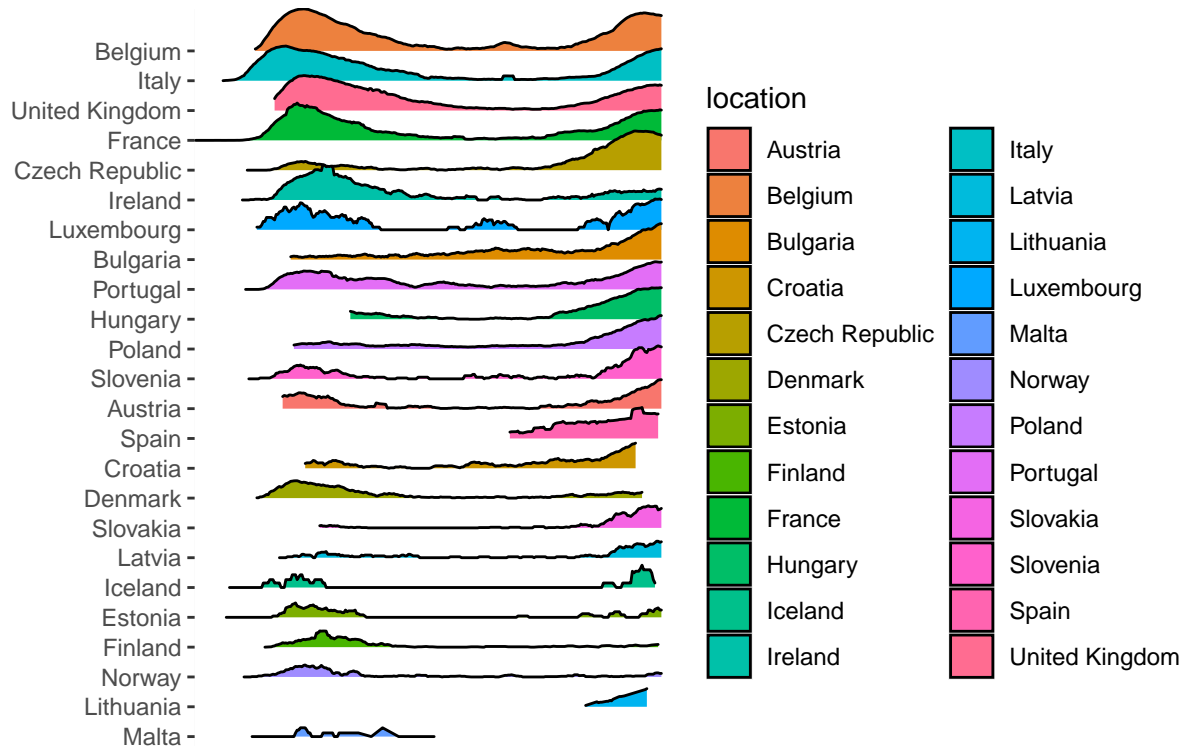
Dane na milion

Wykres pokazuje porównanie wygładzenia (średni kroczącej) śmierci na milion pomiędzy krajami Europy. Należy zaznaczyć, że wysokość wykresu odpowiada nie liczbie śmierci a rzędowi wielkości tej liczby.

```
ggplot( data_for_death_plots,
  aes(x= date,
    y = fct_reorder( location, total_deaths_per_million),
    height = log( new_deaths_smoothed_per_million + 1)/log(10),
    group = location,
    fill = location)
) +
geom_ridgeline( ) +
xlab("") +
ylab("") +
theme(axis.text.x = element_blank(),
  axis.ticks.x = element_blank()
) +
ggtitle("Porównanie log10(śr. kroc. śmierci/mil. + 1) między krajami EU") +
```

```
theme(plot.title = element_text(hjust = 0.2))
```

Porównanie $\log_{10}(\dots \text{kroc.} \dots \text{mierci} / \text{mil.} + 1)$ między krajami EU



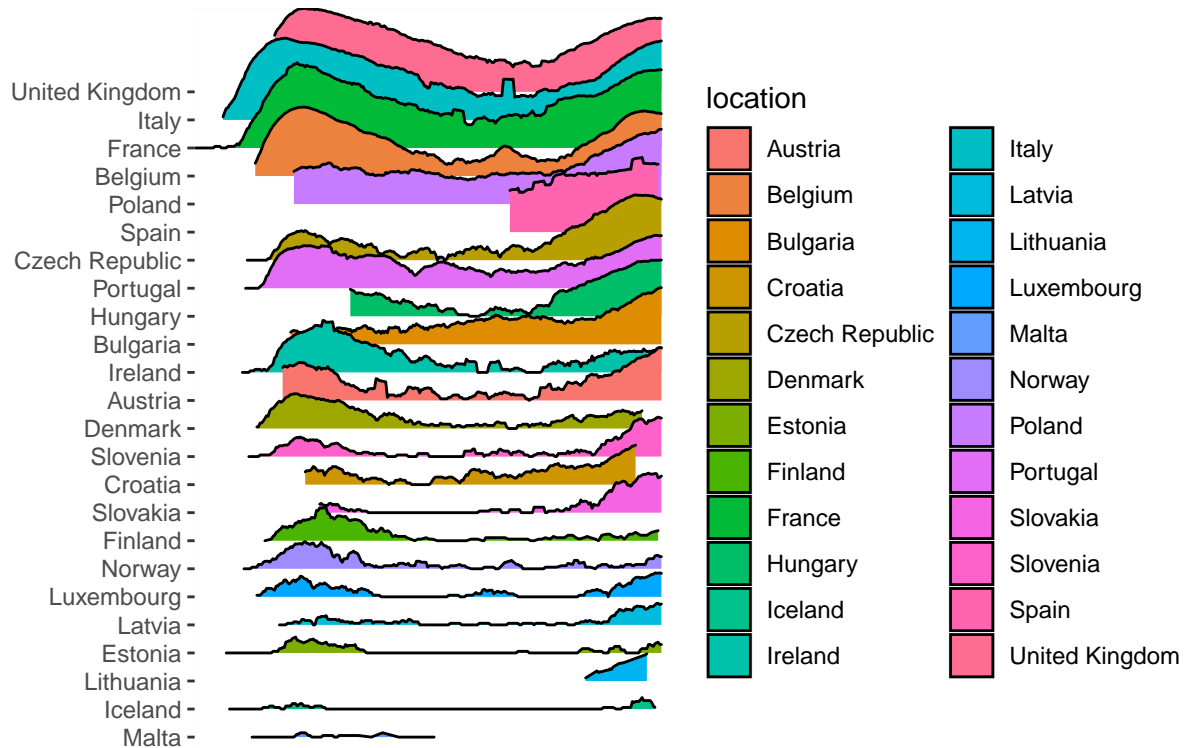
Przy jego pomocy można zauważyć, które kraje EU były najmocniej dotknięte epidemią. Należy zaznaczyć, że dane nie są kompletne, przez co w przypadku Hiszpanii nie widać skali jak mocno została ona dotknięta epidemią.

Dane nieuśrednione względem populacji

Wykres pokazuje porównanie wygładzenia (średni kroczącej) śmierci pomiędzy krajami Europy. Należy zaznaczyć, że wysokość wykresu odpowiada nie liczbie śmierci a rzędowi wielkości tej liczby.

```
ggplot( data_for_death_plots, aes( x= date,
                                   y = fct_reorder( location, total_deaths),
                                   height = log( new_deaths_smoothed + 1 ) / log(10),
                                   group = location,
                                   fill = location)
) +
geom_ridgeline( ) +
xlab("") +
ylab("") +
theme(axis.text.x = element_blank(),
      axis.ticks.x = element_blank())
) +
ggtitle("Porównanie  $\log_{10}(\text{śr. krocz. śmierci} + 1)$  między krajami EU") +
theme(plot.title = element_text(hjust = 0.2))
```

Porównanie $\log_{10}(\dots r. \text{ kroc.} \dots \text{mierci} + 1)$ mi..dzy krajami EU



Niestety wykres ten jest znacznie mniej przejrzysty względem poprzedniego.

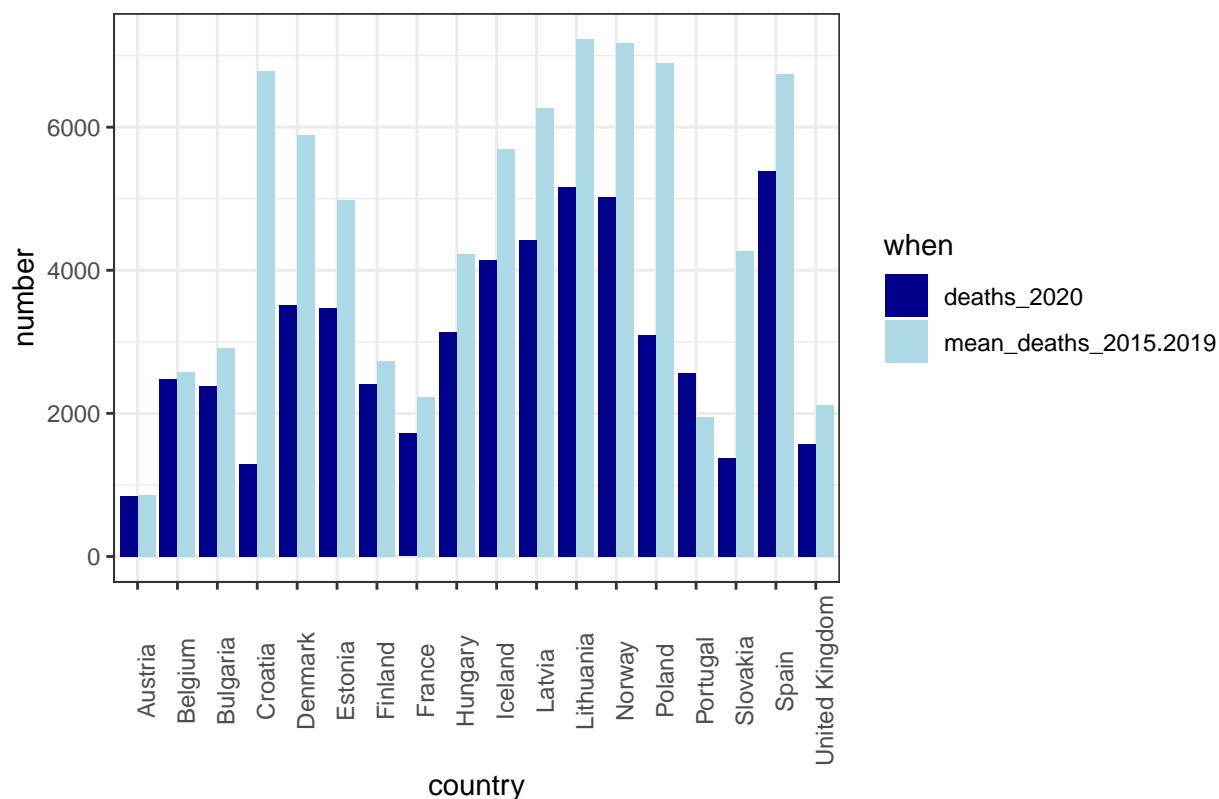
Ogólna liczba zgonów w wybranych europejskich krajach

W celu porównania całkowitej liczby zgonów, użyliśmy danych dostępnych na stronie Eurostat. Zbiór zawierał tygodniową liczbę zgonów w poszczególnych krajach. Wybraliśmy tygodnie 40-44 (październik). Porównaliśmy zgony w roku 2020 z średnią śmierci w tym samym okresie w poprzednich pięciu latach. Wykres przedstawia się następująco:

```
wide = gather(dane_full, "when", "number", "mean_deaths_2015.2019", "deaths_2020", -country)
wide$when <- as.factor(wide$when)
```

```
ggplot(wide, aes(x=country, y=number, fill=when)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_manual(values=c("darkblue", "lightblue")) + theme_bw() +
  ggtitle("Deaths in weeks 40-44 in the selected European countries") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.x = element_text(angle = 90))
```

Deaths in weeks 40–44 in the selected European countries



Wykres wydał nam się podejrzany, zgonów było podejrzenie mało. Po porównaniu danych z Eurostatu z tygodni 40-44 2020 i 2019 dla Polski z analogicznymi danymi z GUSu, okazało się, że dane Eurostatu nie są prawidłowe. Przykładowo: Według danych z GUS z tygodni 40-44 2020 liczba zgonów w Polsce: 53759 Według Eurostatu: 3102

Analogiczna sytuacja jest dla innych krajów.