

Praca domowa 2

Krystian Okręgliński

20 03 2021

Cel zadania

Zadaniem pracy domowej było znalezienie wadliwego lub źle wykonanego wykresu/tabelki oraz pokazanie propozycji jego naprawienia.

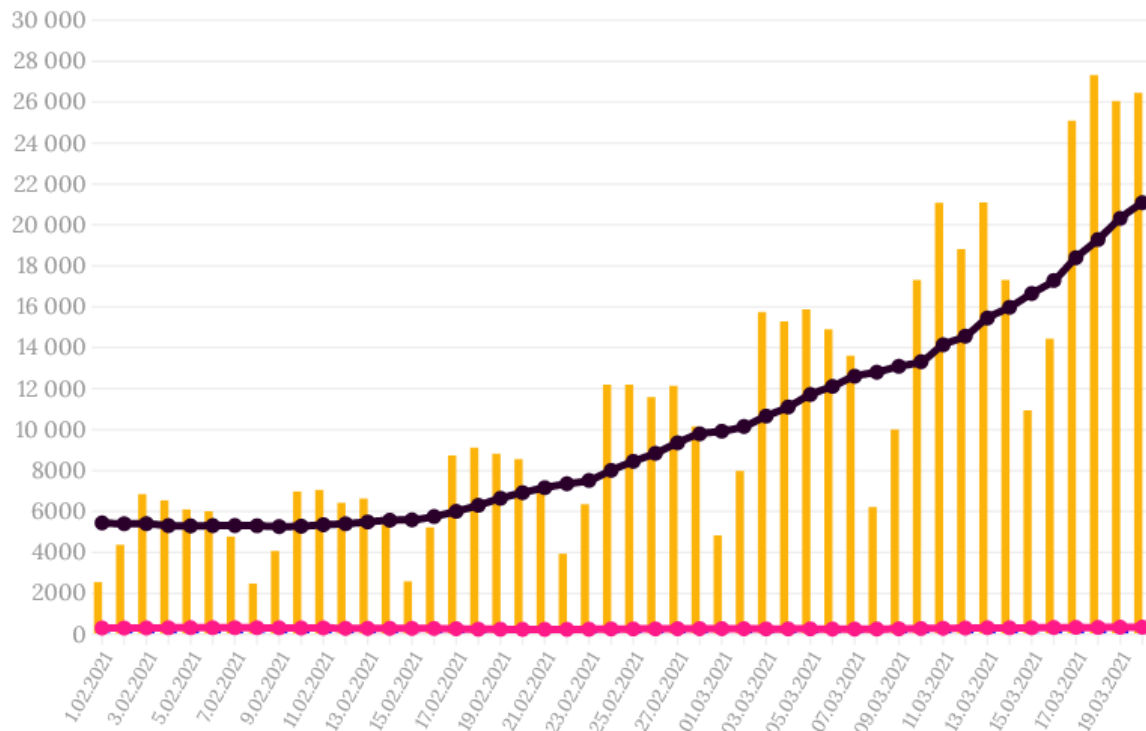
Wadliwy wykres

Do tego zadania wybrałem wykres ze strony <https://fakty.tvn24.pl/ogladaj-online,60/koronawirus-w-polsce-w-szpitalach-coraz-wiecej-mlodych-pacjentow,1053483.html>:

Dzienny bilans zakażeń i zgonów oraz średnia z siedmiu ostatnich dni w Polsce

Wykres prezentuje dane od początku lutego

■ Średnia zakażeń z siedmiu dni ■ Średnia zgonów z siedmiu dni
■ Nowe potwierdzone przypadki zakażenia ■ Nowe przypadki śmiertelne



Źródło: Ministerstwo Zdrowia

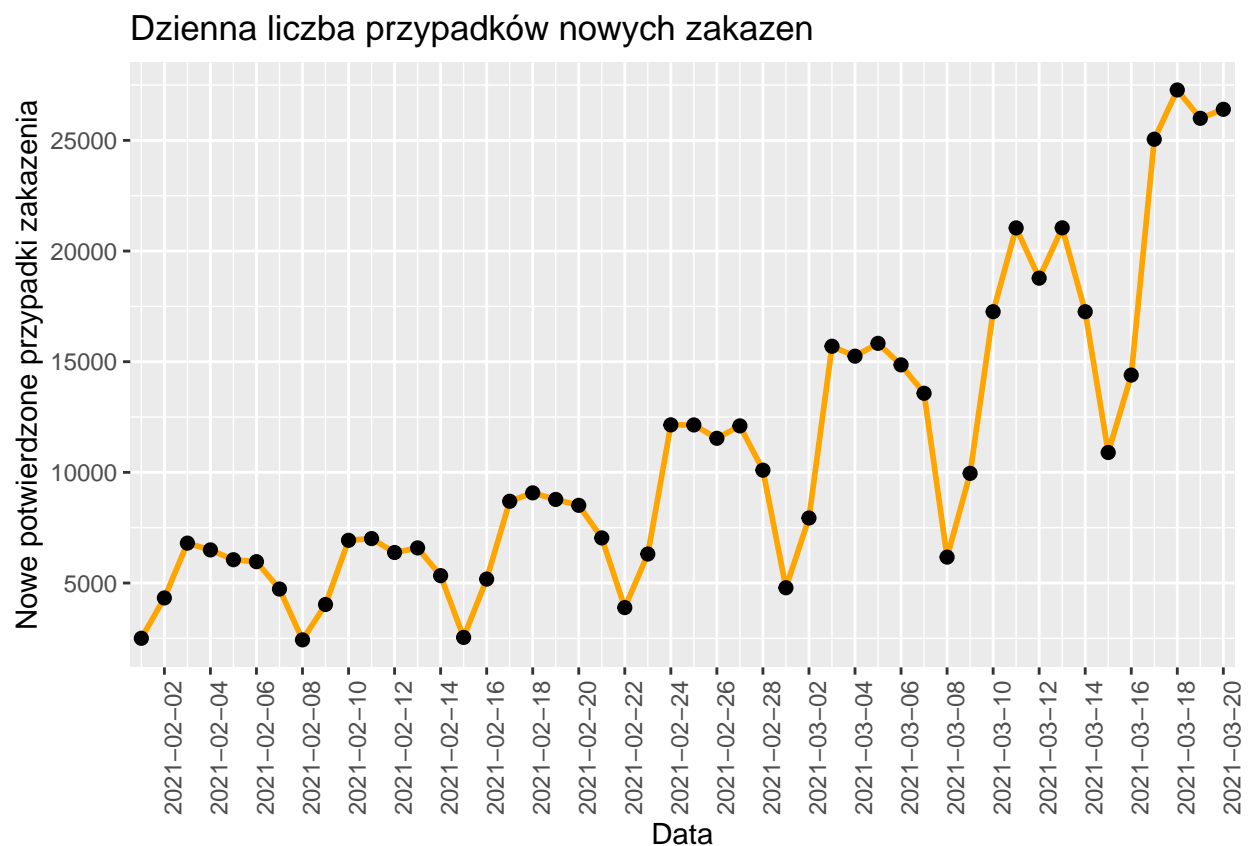
Widzimy, że na wykres ten nałożone są cztery dane, z czego informacje o “Średniej zgonów z siedmiu dni” oraz

“Nowe śmiertelne przypadki” są znacząco mniejsze od “Średniej zakażeń z siedmiu dni” i “Nowe potwierdzone przypadki zakażenia”, co powoduje zupełną nieczytelność pierwszych dwóch. Poza tym sama koncepcja “średniej ... z siedmiu dni” jest niezupełnie jasna i nie wiadomo, do jakiego konkretnie okresu się odnosi.

Propozycja poprawnego wykresu

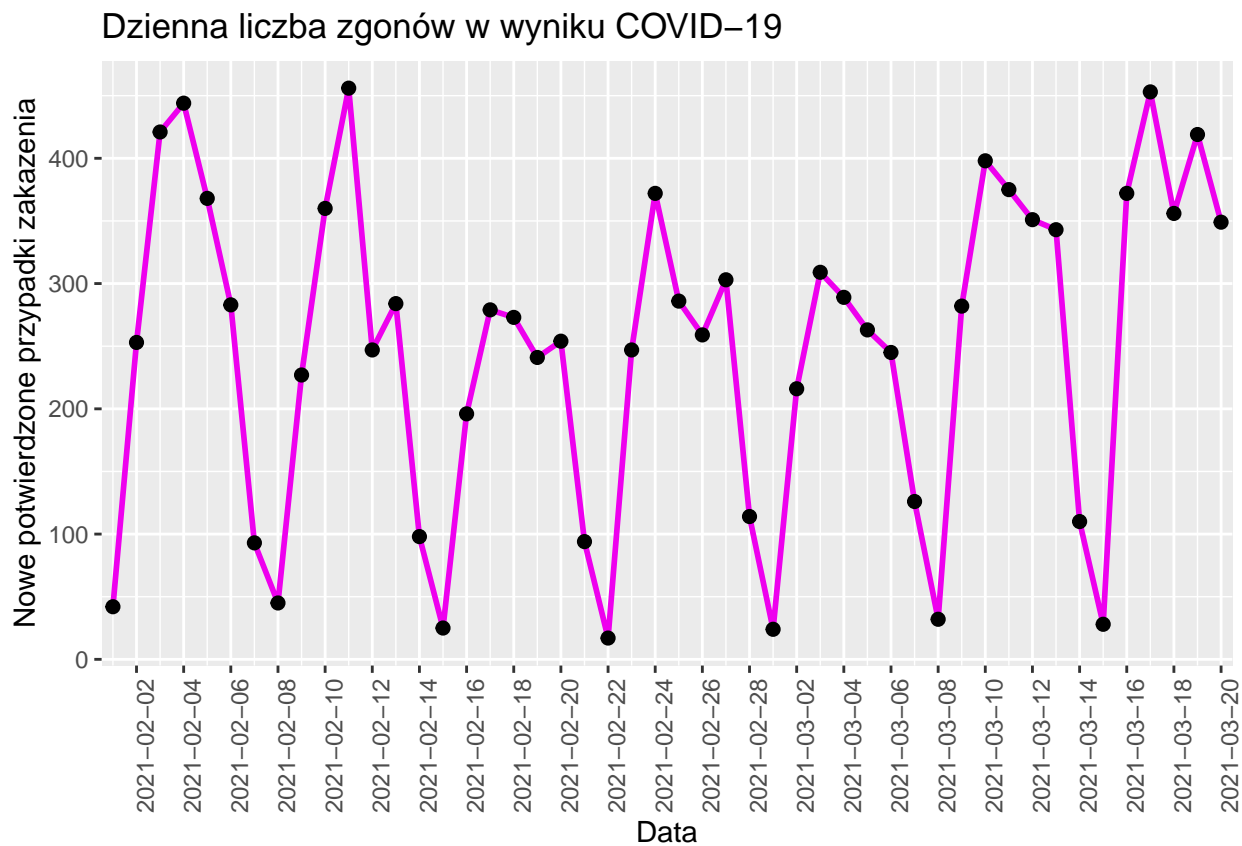
Zamiast zamieszczenia praktycznie czterech wykresów na raz, zdecydowałem się na oddzielne ich pokazanie oraz całkowitą rezygnację z danych “Średnia zgonów z siedmiu dni” i “Średnia zakażeń z siedmiu dni”. Do tworzenia wykresu będę korzystał z obrobionych już przeze mnie danych z oficjalnej strony Ministerstwa Zdrowia. Zdecydowałem także, że zamiast prezentacji danych w postaci wykresu słupkowego, przedstawię je jako wykres liniowy z wyszczególnionymi punktami.

```
dane <- fread("./dane.csv")
dane$Data <- as.Date(dane$Data)
ggplot(dane, aes(x=Data, y=liczba_przypadkow))+
  geom_line(size=1, color = "orange")+
  geom_point(size=2)+
  labs(title = "Dzienna liczba przypadków nowych zakażeń",
       x = "Data",
       y = "Nowe potwierdzone przypadki zakażenia")+
  theme(axis.text.x = element_text(angle = 90))+
  scale_x_date(breaks = breaks_width("2 days"), expand = c(0.01,0.01))+
  scale_y_continuous(breaks = breaks_width(5000))
```



```
ggplot(dane, aes(x=Data, y=zgony))+
  geom_line(size=1, color = "magenta2")+
  geom_point(size=2)+
```

```
theme(axis.text.x = element_text(angle = 90))+
scale_x_date(breaks = breaks_width("2 days"), expand = c(0.01,0.01))+
scale_y_continuous(breaks = breaks_width(100))+
labs(title = "Dzienna liczba zgonów w wyniku COVID-19",
x = "Data",
y = "Nowe potwierdzone przypadki zakażenia")
```



Uzasadnienie

Uważam, że moje przedstawienie danych w postaci dwóch wykresów jest lepsze, ponieważ została poprawiona czytelność informacji, w szczególności można rzeczywiście odczytać dane o liczbie przypadków śmiertelnych. Także, ponieważ dane o "średnich" mają niejasną metodę wyznaczenia, uważam, że nieprezentowanie ich jest dobrą decyzją.