

[TWD] Praca domowa 2

Kacper Kurowski

18 10 2020

Problematyczny wykres

Poniższy wykres znalazłem na stronie glos24.pl (dostęp 17.10.2020). Przedstawia on liczbę potwierdzonych zakażeń wirusem SARS-COV-2 w Polsce między 24 września a 15 października.

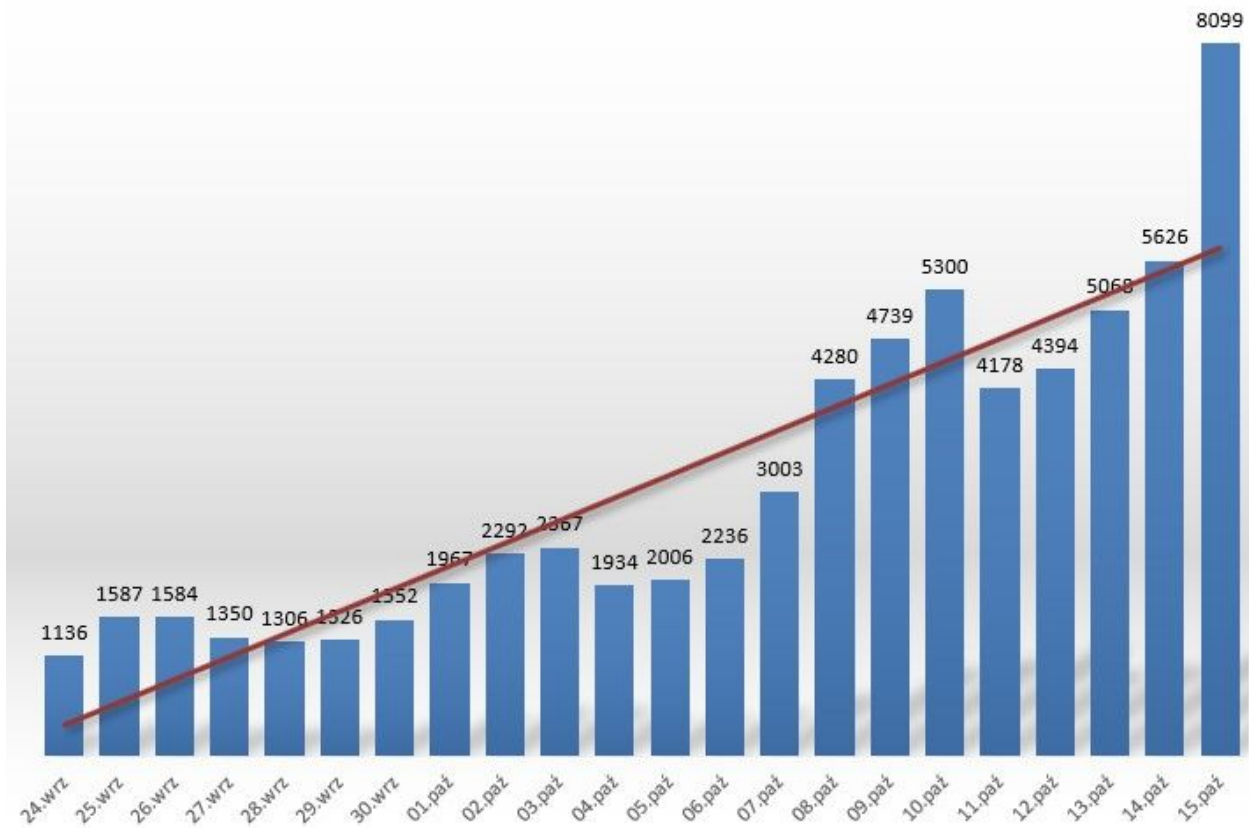


Figure 1: Wykres liczby zakażeń

Wykres ten nie jest opatrzony ani tytułem, ani osią pionową, na podstawie których byłoby możliwe określenie jego celu. Powyżej słupków umieszczone zostały dokładne liczby zakażeń potwierdzonych danego dnia - nie wynika to jednak ani z wykresu, ani z treści artykułu - wykres ten został do niego “doczepiony”.

Możnaby się zastanowić, czy podanie dokładnych wartości jest tutaj konieczne - wydaje się, że celem wykresu jest przedstawienie tendencji wzrostowej oraz skali a nie precyzyjnych danych. Liczby te często są przesłonięte przez czerwoną kreskę, której cel jest nieznany. Nie wiadomo, czy przedstawia ona regresję liniową, czy może trend wzrostowy widziany “na oko” - ten widać z samych słupków.

Kolejnym zbytecznym elementem graficznym są cienie za słupkami - niczemu one nie służą. Da się również zauważyć gradient ściemniający dolną część wykresu - kolejny niepotrzebny element graficzny. Także czerwona kreska rzuca cień na resztę wykresu.

Wczytanie danych

```
zakazenia <- c(1136, 1587, 1584, 1350, 1306, 1526, 1552, 1967, 2292, 2367, 1934,
               2006, 2236, 3003, 4280, 4739, 5300, 4178, 4394, 5068, 5626, 8099)
daty <- as.Date( c('24.09', '25.09', '26.09', '27.09', '28.09', '29.09', '30.09', '01.10',
                  '02.10', '03.10', '04.10', '05.10', '06.10', '07.10', '08.10', '09.10',
                  '10.10', '11.10', '12.10', '13.10', '14.10', '15.10'),
               tryFormats = '%d.%m')
dane <- data.frame(daty, zakazenia)
```

Próba poprawienia wykresu

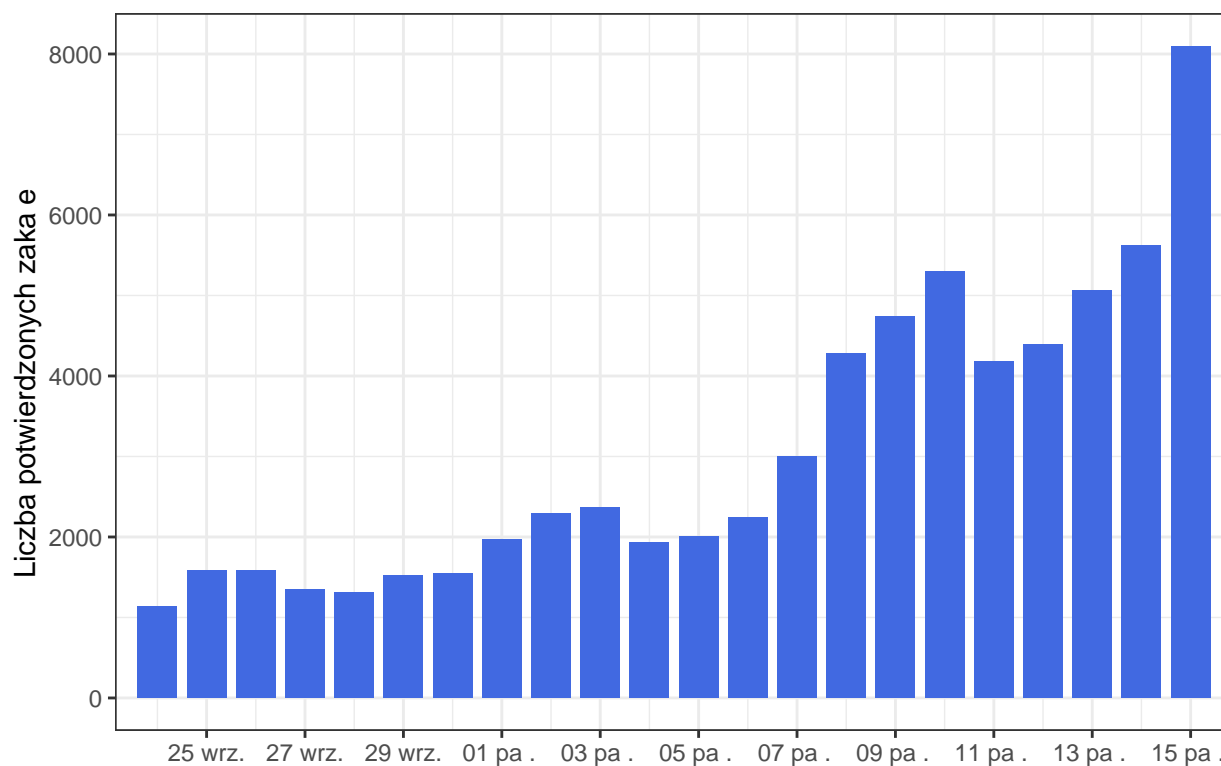
W celu poprawienia powyższego wykresu dokonałem następujących zmian:

- Dodałem tytuł wykresu
- Dodałem oś pionową i opis co ma reprezentować
- Usunąłem liczby powyżej słupków
- Usunąłem czerwoną kreskę
- Usunąłem cienie za słupkami i gradient ściemniający dół wykresu
- Na osi x daty sprecyzowane są co drugi dzień - ułatwia to ich odczytanie

```
library(ggplot2)

ggplot( dane, aes(x = daty, y = zakazenia, width=0.8) ) +
  xlab('') +
  ylab('Liczba potwierdzonych zakażeń') +
  ggtitle('Potwierdzone zakażenia koronawirusem w Polsce 24 wrz. - 15 paź. 2020') +
  geom_col(fill = '#4169E1') +
  scale_x_date( date_labels = '%d %b.', date_breaks = '2 days', expand =c(0.02, 0.02)) +
  theme_bw()
```

Potwierdzone zakażenia koronawirusem w Polsce 24 wrz. – 15 pa . 2020



Utworzony w ten sposób wykres jest przejrzysty oraz wiadomo jakie dane są reprezentowane. Rzutem oka da się stwierdzić trend wzrostowy - bez czerwonej linii. Zamiast dokładnych wartości mamy ich przybliżenia - zamiast wczytywać się w liczby na każdym słupku wystarczy raz spojrzeć na oś pionową, by stwierdzić, że zakażenia liczymy już w tysiącach.