

# Hoang Laura: praca domowa nr.2

autor: Hoang Laura

01/11/2021

## Wstęp

Celem tej pracy, było stworzenie własnej wersji pewnej wizualizacji, udostępnionej publicznie, w przeciągu ostatnich 2 tygodni.

Jako źródło, postanowiłam wybrać stronę <https://www.gov.pl/web/koronawirus/wykaz-zarazen-koronawirusem-sars-cov-2>, w której mamy dostępne dane dotyczące liczby zakażeń koronawirusem z każdego dnia. Zdecydowałam się na analizę dnia 01.11.2021.

Importujemy biblioteki:

```
library(dplyr)
library(ggplot2)
```

Wczytujemy dane:

```
covid_original <- read.csv("akutalne_dane_wojewodztwa.csv",
                           sep=";",
                           stringsAsFactors=TRUE)

covid <- covid_original[-c(1),]
rownames(covid) <- 1:16
covid <- arrange(covid,-liczba_przypadkow)

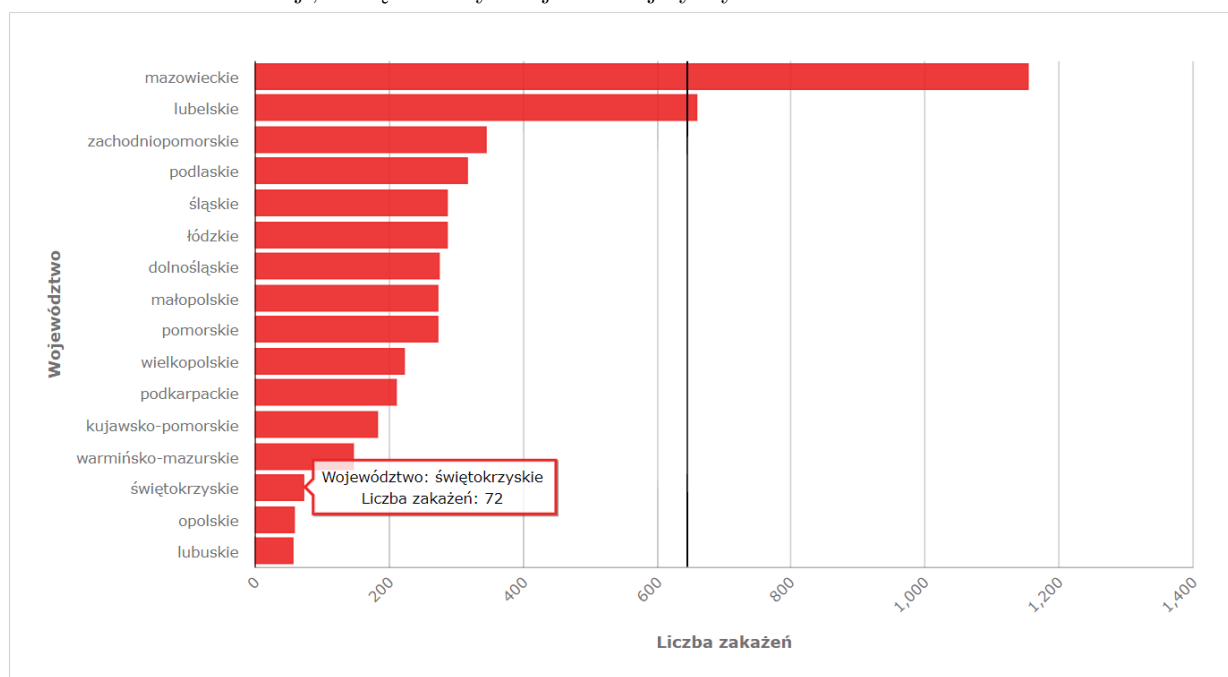
head(covid)
```

```
##      wojewodztwo  liczba_przypadkow  liczba_na_10_tys_mieszkancow  zgony
## 1      mazowieckie             1154                2.13          6
## 2      lubelskie             658                3.13          1
## 3 zachodniopomorskie             344                2.03          0
## 4      podlaskie             317                2.69          0
## 5      śląskie             287                0.64          0
## 6      łódzkie             286                1.17          1
##  zgony_w_wyniku_covid_bez_chorob_wspolistniejacych
## 1                                5
## 2                                0
## 3                                0
## 4                                0
## 5                                0
## 6                                0
##  zgony_w_wyniku_covid_i_chorob_wspolistniejacych  liczba_zleceń_poz
```

## 1	1	95	
## 2	1	66	
## 3	0	9	
## 4	0	9	
## 5	0	50	
## 6	1	24	
##	liczba_ozdrowiencow	liczba_osob_objetych_kwarantanna	liczba_wykonanych_testow
## 1	1308	62013	5332
## 2	1402	33615	2181
## 3	302	15426	1627
## 4	587	20041	1093
## 5	277	14074	2721
## 6	285	16508	1902
##	liczba_testow_z_wynikiem_pozytywnym	liczba_testow_z_wynikiem_negatywnym	
## 1	1285	3980	
## 2	716	1442	
## 3	393	1220	
## 4	348	726	
## 5	327	2380	
## 6	312	1582	
##	liczba_pozostalych_testow	teryt	stan_rekordu_na
## 1	67	t14	2021-10-31
## 2	23	t06	2021-10-31
## 3	14	t32	2021-10-31
## 4	19	t20	2021-10-31
## 5	14	t24	2021-10-31
## 6	8	t10	2021-10-31

## Oryginalna wizualizacja

Pierwowzorem wizualizacji, którą utworzyłam jest niniejszy wykres:

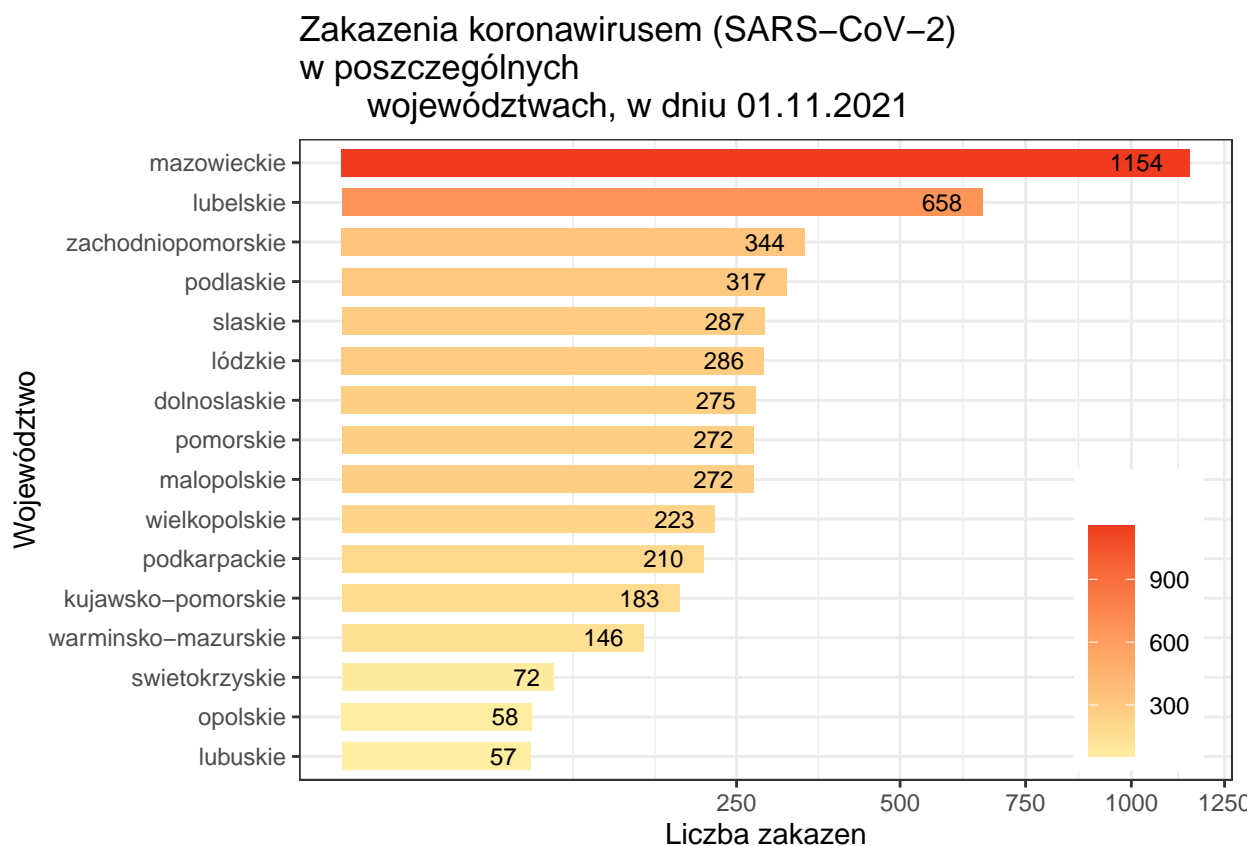


## Moja wizualizacja

Kod i utworzona wizualizacja:

```
p_col <- covid %>%
  ggplot( aes(x = liczba_przypadkow, y = reorder(województwo, liczba_przypadkow),
            fill = liczba_przypadkow) ) +
  geom_col(width=.7) +
  scale_fill_gradient(low = "#ffeda0", high = "#f03b20") +
  geom_text( aes(label = liczba_przypadkow), hjust = 1.5, size = 3) +
  xlab("Liczba zakażeń") +
  ylab("Województwo") +
  theme_bw() +
  scale_x_continuous(trans = "sqrt") +
  labs(title = "Zakażenia koronawirusem (SARS-CoV-2) \nw poszczególnych
            województwach, w dniu 01.11.2021", fill = "") +
  theme(legend.position = c(0.9, 0.25))

p_col
```



## Podsumowanie

W mojej wizualizacji nie zmieniałam typu wykresu, ale poprawiłam jego czytelność i estetykę.

Ponieważ występuje duży uskoc między największą i najmniejszą wartością przy **liczbie zakażeń**, oryginalny wykres ma pustą przestrzeń w prawym dolnym rogu, przez co nie wygląda on estetycznie.

Dlatego, przeskalowałam oś X-ów. Zdecydowałam się o przeskalowanie o pierwiastek kwadratowy, aby zapobiec temu problemowi, ale zarazem nie psuło to czytelności wykresu (jak by się działo, gdybym dokonała skalowania o logarytm).

Oprócz tego, dodałam etykiety z wartościami na końcu każdego słupka i ustawiłam kolorystykę gradientową (wraz z legendą), która zmienia się względem wartości. Uważam, że jest to lepsze rozwiązanie niż interaktywne elementy oryginału, gdyż w tamtym wypadku etykiety i pionowa linia przeszkadzały w odczytywaniu danych i zasłaniały pozostałe słupki.