

Praca domowa numer 2

Tymoteusz Barciński

3 listopada 2021

1. Wprowadzenie

Celem pracy domowej było przygotowanie wizualizacji utworzonej z wykorzystaniem pakietu ggplot2 na podstawie wizualizacji dostępnej w prasie, telewizji lub internecie, tak aby była ona czytelniejsza.

2. Oryginalny wykres

Liczba mieszkańców województwa na jednego zakażonego w tym województwie

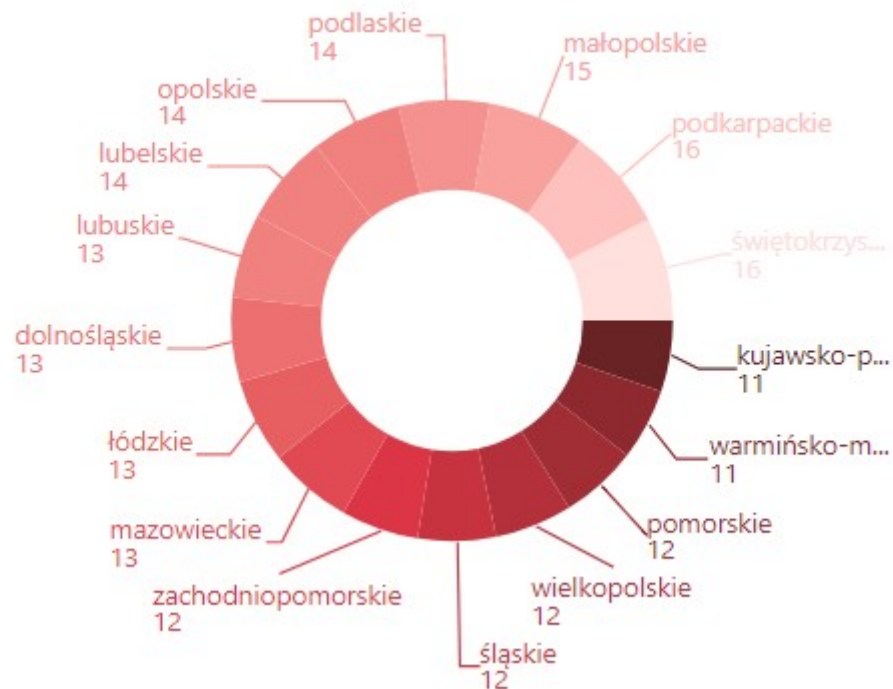


Figure 1: źródło: koronawirusunas.pl, dane z dnia 3 listopada 2021

3. Kod oraz poprawiony wykres

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

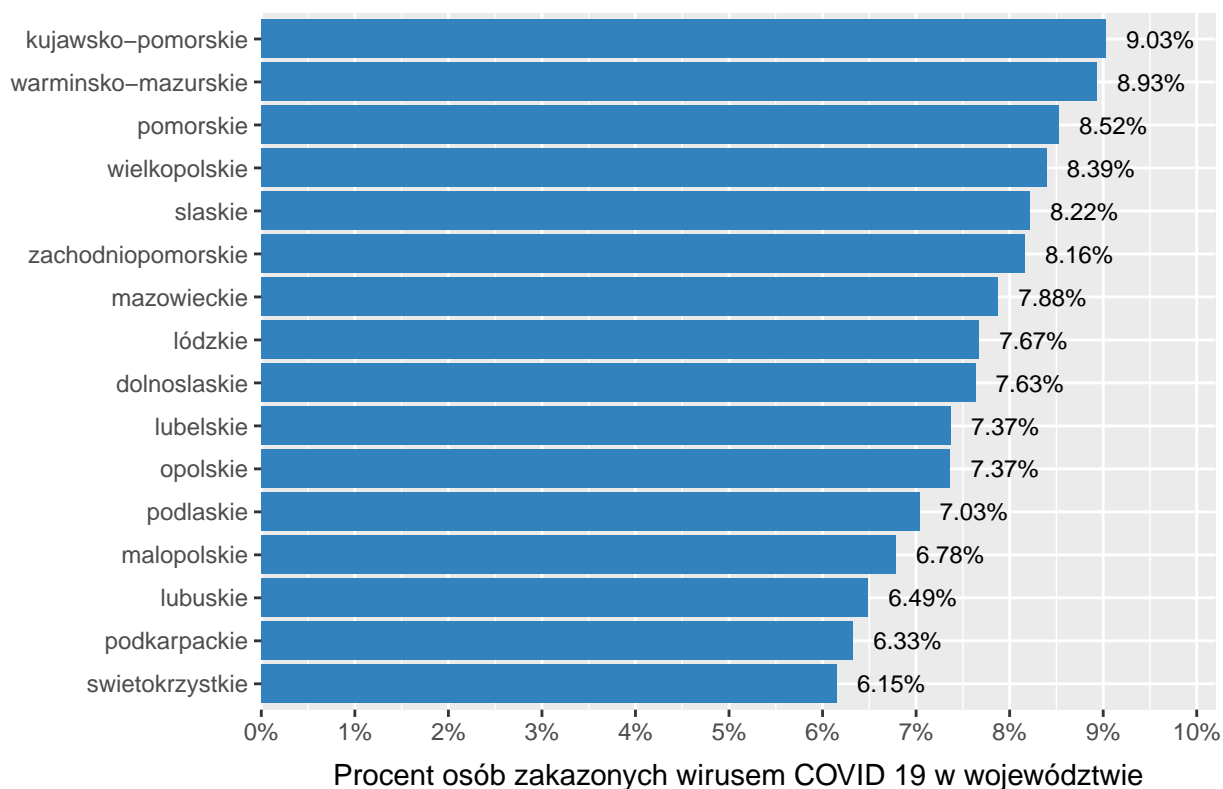
województwa <- c("podlaskie", "warmińsko-mazurskie", "lubuskie", "zachodniopomorskie",
                 "lubelskie", "świętokrzyskie", "podkarpackie", "opolskie",
                 "wielkopolskie", "łódzkie", "kujawsko-pomorskie", "pomorskie",
                 "dolnośląskie", "mazowieckie", "małopolskie", "śląskie")
liczba_zakazonych_na_100km2 <- c(408, 527, 545, 604, 619, 646, 753, 766, 982, 1036, 1038,
                                1091, 1107, 1205, 1526, 3007)
gestosc_zaludnienia_na_1km2 <- c(58, 59, 84, 74, 84, 105, 119, 104, 117, 135, 115, 128,
                                145, 153, 225, 366)

dane <- data.frame(województwa, liczba_zakazonych_na_100km2, gestosc_zaludnienia_na_1km2)

dane <- dane %>%
  mutate(procent_zakazonych = (liczba_zakazonych_na_100km2 /
                               (gestosc_zaludnienia_na_1km2*100)) * 100)

ggplot(dane, aes(x = procent_zakazonych, y = fct_reorder(factor(województwa),
                                                         procent_zakazonych))) +
  geom_col(fill = "#3182bd") +
  scale_x_continuous(labels = function(x) paste0(x, "%"), breaks = seq(0, 11, by = 1),
                    expand = expansion(mult = c(0, 0.13))) +
  labs(x = "Procent osób zakażonych wirusem COVID 19 w województwie", y = element_blank(),
       title = "Procent osób zakażonych wirusem COVID 19 w poszczególnych województwach") +
  geom_text(aes(label = paste0(round(procent_zakazonych, 2), "%")),
            hjust = -0.3, size = 3.3) +
  theme(axis.text=element_text(size=9),
        axis.title=element_text(size=11),
        axis.title.x = element_text(margin = margin(t = 7, r = 0, b = 0, l = 0)),
        plot.title = element_text(hjust = 1,
                                   margin = margin(t = 0, r = 10, b = 10, l = 0)),
        title = element_text(size = 10.3))
```

Procent osób zakażonych wirusem COVID 19 w poszczególnych województwach



4. Uzasadnienie

W oryginalnej wizualizacji został użyty wykres pierścieniowy, który nie jest najlepszym sposobem prezentowania danych. Został on zamieniony na wykres słupkowy, aby poprawić czytelność. Ponadto, uważam, że uwzględnienie gęstości zaludnienia w poszczególnych województwach¹ i zaprezentowanie danych w procentach, zwiększa klarowność wykresu. Pierwotna wizualizacja miała również nieumiejętnie dobraną skalę kolorów - zastosowanie gradientu kolorów było zbyt trudne. Znacząco utrudniało to odczytanie niektórych nazw województw. Poprawiony wykres uwzględnia wartość dla województw za pomocą ich uszeregowania i nie wprowadza niepotrzebnych kolorów, zwiększając przejrzystość wizualizacji.

¹GUS - Powierzchnia i ludność w przekroju terytorialnym w 2020 roku