

Raport - HW2

Łukasz Tomaszewski

1. Poprawiana wizualizacja

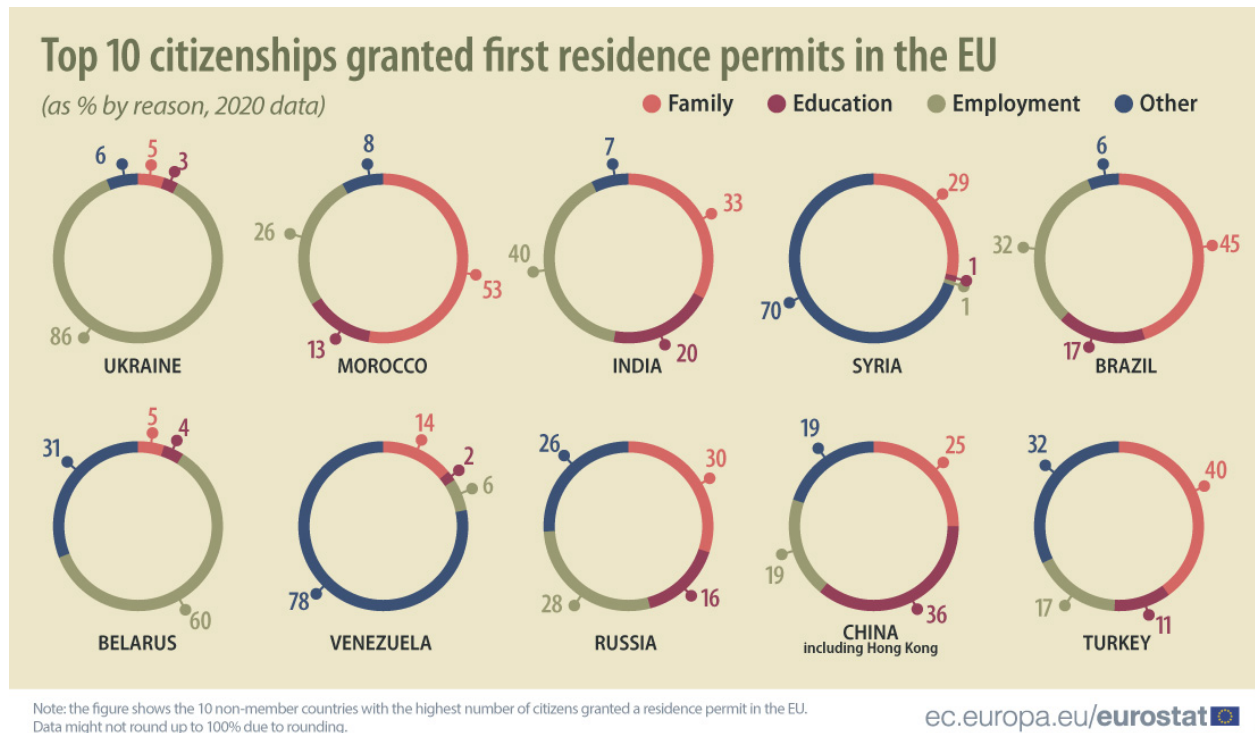


Figure 1: Źródło wizualizacji: <https://next.gazeta.pl/next/7,151003,27676154,migranci-w-polsce-wydajemy-najwiecej-pozwolen-ze-wszystkich.html>

2. Własna wizualizacja

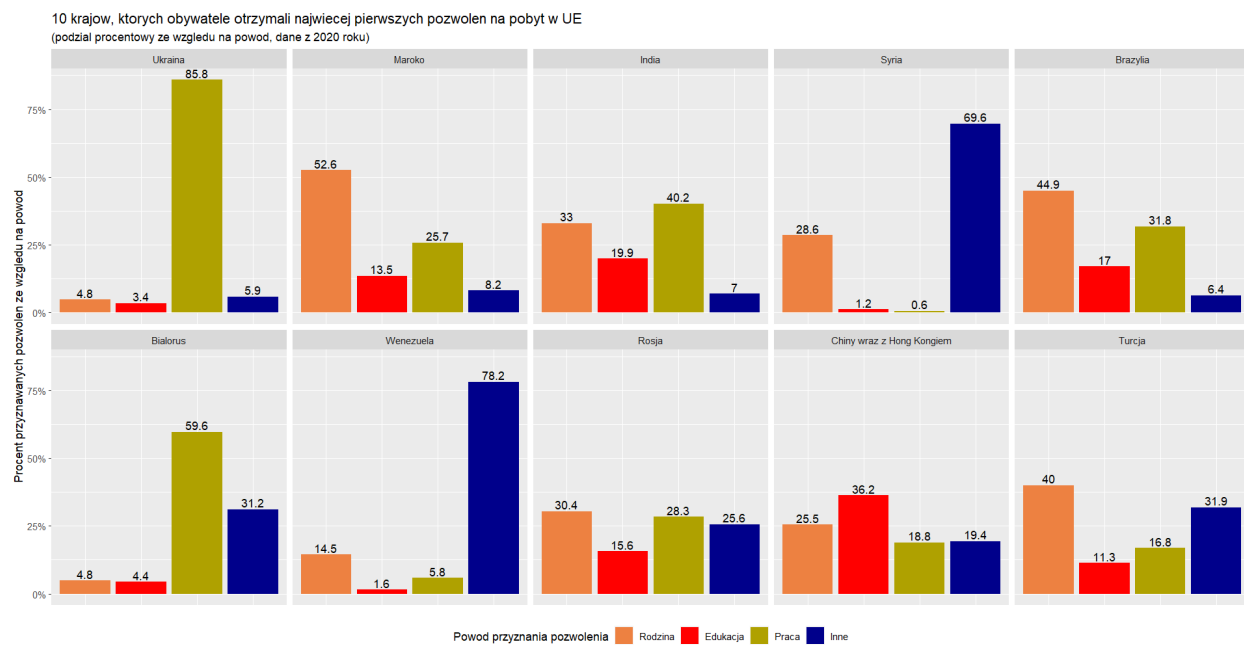


Figure 2: Poprawiona wizualizacja

3. Kod

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

df <- read.csv("estat_migr_resfirst_filtered_en.csv")

df <- df %>% select(reason,citizen,OBS_VALUE) %>% group_by(citizen) %>%
  mutate(Procent = OBS_VALUE/max(OBS_VALUE), TOTAL = max(OBS_VALUE)) %>%
  filter(reason != "TOTAL") %>% arrange(-TOTAL)
df[2] <- rep(c("Ukraina","Maroko","India","Syria","Brazylia",
              "Białorus","Wenezuela","Rosja","Chiny wraz z Hong Kongiem",
              "Turcja"), each = 4)
df[1]<- rep(c("Edukacja","Rodzina","Praca","Inne"), 10)
df$reason <- factor(df$reason, levels = c("Rodzina","Edukacja","Praca","Inne"))

plot <- ggplot(df,aes(x = reason, y = Procent, fill = reason)) +
  geom_col(position = "dodge") +
  facet_wrap(reorder(df$citizen, -(df$TOTAL)), nrow = 2) +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(), axis.ticks.x=element_blank()) +
  ylab("Procent przyznawanych pozwoleń ze względu na powód")+
  labs(fill = "Powód przyznania pozwolenia",
       title = "10 krajów, których obywatele
       otrzymali najwięcej pierwszych pozwoleń na pobyt w UE",
       subtitle = "(podział procentowy ze względu na powód, dane z 2020 roku)") +
  theme(legend.position="bottom") +
  geom_text(aes(label=round(Procent*100,1)),
            position=position_dodge(width=0.9), vjust=-0.25) +
  scale_fill_manual(values=c("#ED8141","red","#AFA100","darkblue")) +
  scale_y_continuous(labels = scales::percent_format())
plot
```

4. Uzasadnienie

Dane na oryginalnym wykresie zostały przedstawione za pomocą wykresów pierścieniowych, których odczytanie może być problematyczne. Odbiorca nie jest w stanie porównać w szybki sposób danych na nich zawartych, zwłaszcza, że te nie są posegregowane względem zaznaczonego obszaru. Dodatkowo zamalowana część nie odpowiada dokładnym danym, a danym zaokrąglonym. W swojej wizualizacji postanowiłem przedstawić dane za pomocą wykresów słupkowych, które lepiej przedstawiają zawarte na nich informacje. Do wykonania wizualizacji wykorzystałem surowe dane. Dzięki temu wykresy są bardziej dokładne. Tekst mówiący o wartości danego słupka też jest dokładniejszy niż w oryginale, ponieważ został on zaokrąglony w mniejszym stopniu.