

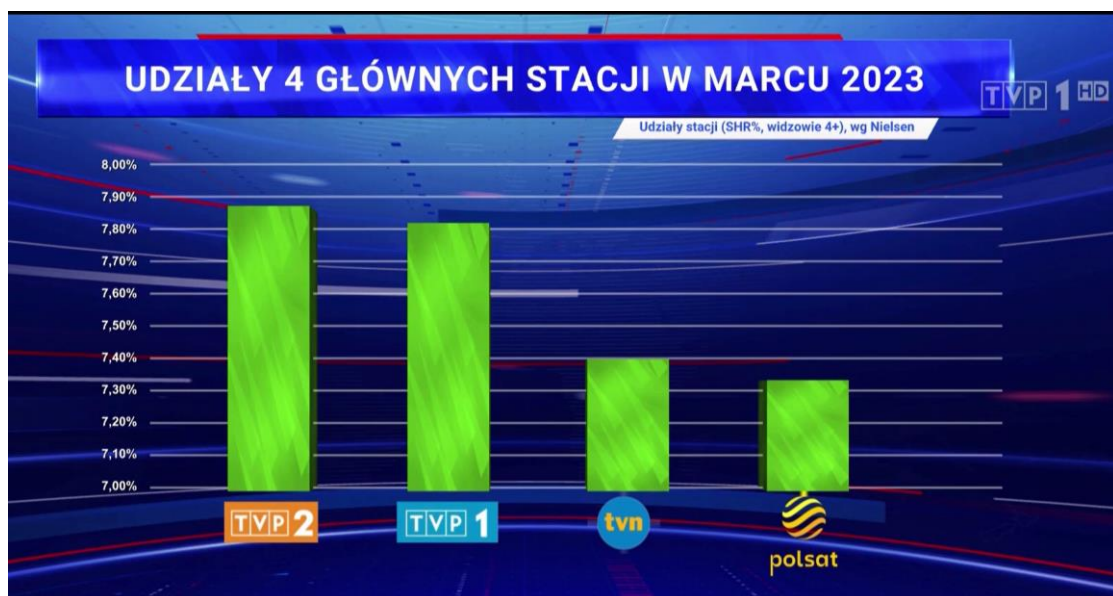
Wstęp do eksploracji danych

Praca domowa 2

Oliwia Trzcińska

Celem tego raportu jest zaprezentowanie wyników pracy domowej 2. Na przykładzie znalezionej wykresu pokazane zostaną błędy popełnione podczas jego tworzenia oraz sposób ich naprawienia.

Wizualizacja, którą chciałabym poprawić, pochodzi ze strony internetowej: <https://wiadomosci.tvp.pl/68490132/widzowie-wybieraja-telewizje-polska?fbclid=IwAR07omdYN8xCcEmthbTSPM6wmDVNEgSScWC5CzNr86a1CLqWGLBoepAtDWo>



Aby stworzyć wykres prezentujący te same dane, ale w sposób bardziej poprawny, użyłam pakietu ggplot2. Poniżej znajduje się kod, który wygenerował nową wizualizację:

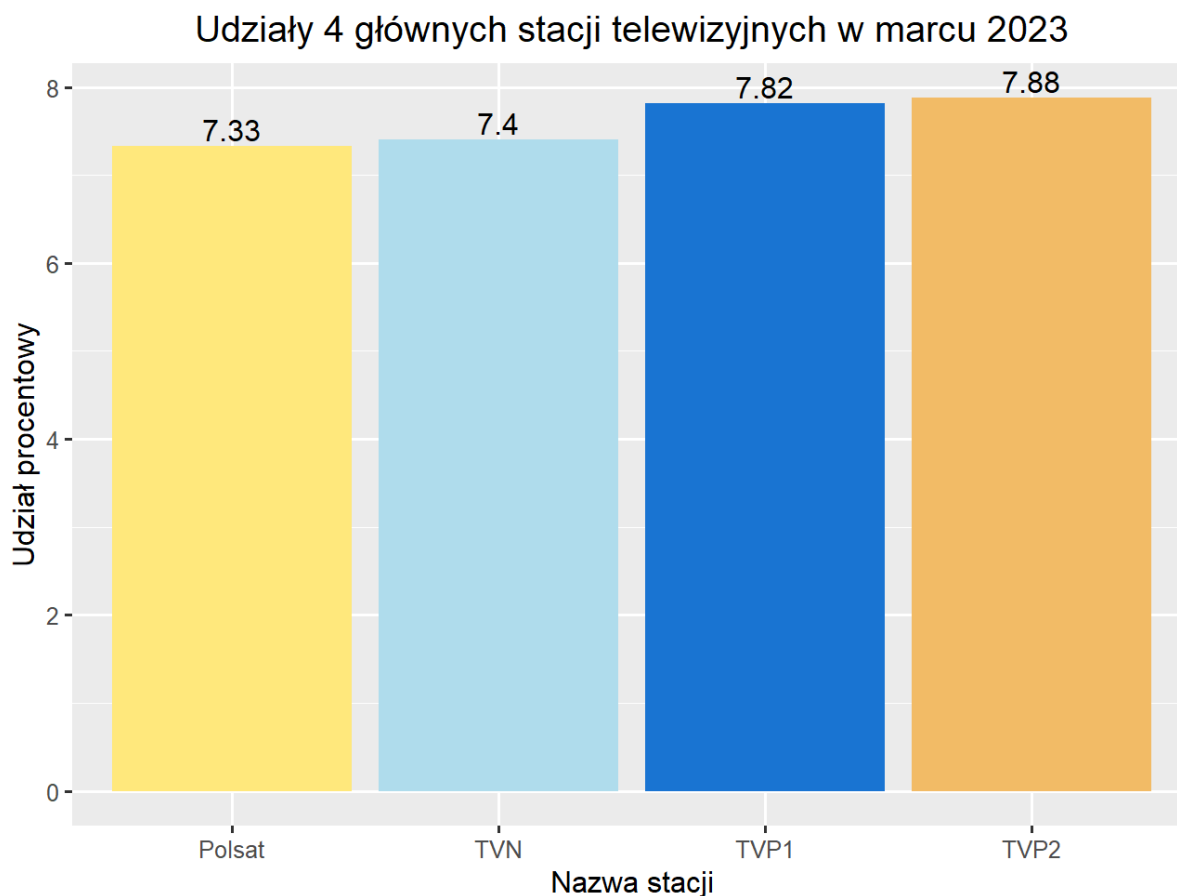
```
library(ggplot2)

df <- data.frame(name = c('TVP2', 'TVP1', 'TVN', 'Polsat'),
                 percentage = c(7.88, 7.82, 7.40, 7.33))

ggplot(df, aes(x = name, y = percentage)) +
  geom_col(fill = c("#F2BB66", "#1974D2", "#AFDCEC", "#FFE87C")) +
  labs(title = "Udziały 4 głównych stacji telewizyjnych w marcu 2023",
       x = "Nazwa stacji",
       y = "Udział procentowy") +
  geom_text(aes(label=percentage),
            position=position_dodge(width=0.9),
```

```
vjust=-0.25) +  
  theme(plot.title = element_text(hjust = 0.5))
```

Powstał dzięki niemu następujący wykres:



Poprawiona wizualizacja jest lepsza niż oryginalna, ponieważ naprawia podstawowy błąd: na pierwszym wykresie oś Y jest wyskalowana od 7%. Dzięki temu różnica między wartościami dla konkurencyjnych stacji telewizyjnych wygląda na dużo większą niż w rzeczywistości jest. Skala na nowym wykresie zaczyna się od 0%, co pozwala zauważyć, że różnice między kolejnymi wartościami są znikome. Dodatkowo na każdym słupku została dodana wartość liczbową, aby ułatwić odczytywanie wykresu.