

Techniki Wizualizacji Danych

Raport z pracy domowej nr 2

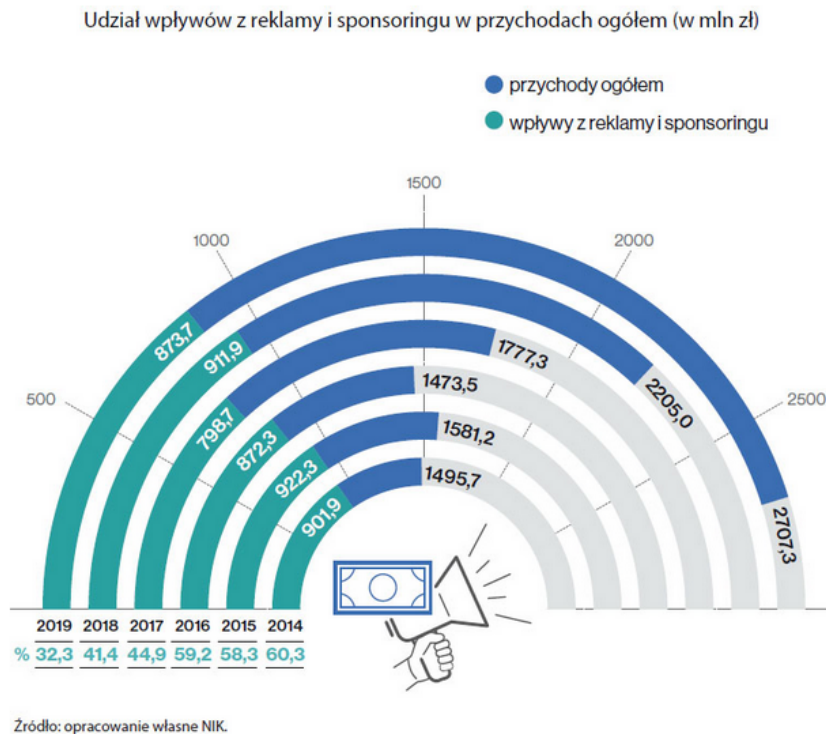
Jakub Piwko

2 listopada 2021

Wprowadzenie

Niniejszy raport został opracowany w ramach przedmiotu Techniki Wizualizacji Danych na kierunku Inżynieria i Analiza Danych w semestrze zimowym 2021-2022. Jego celem jest zaprezentowanie wyników pracy nad udoskonaleniem odszukanego w prasie, internecie lub telewizji wykresu, używając pakietu ggplot2 w języku R.

Wykres do poprawy



Rysunek 1: Znaleziony wykres

Powyższy wykres i dane, na bazie których został tworzony, pochodzą z [z raportu Najwyższej Izby Kontroli](#) dotyczącego gospodarki majątkowej Telewizji Polskiej S.A. opublikowanego na oficjalnej stronie instytucji 15 października bieżącego roku. Przedstawia on udział wpływów z reklam w przychodach całkowitych w ostatnich latach. Można powiedzieć, że mamy tu do czynienia z kombinacją wykresu słupkowego i kołowego. Przez niecodzienny kształt, na pierwszy rzut oka trudno jest odczytać dane z wykresu. Słupki z niezrozumiałego powodu są wygięte w kształt półokrągły. Dodatkowo, nieintuicyjnie, wartości słupków wyrażane są nie przez ich długości, ale przez kąt, do którego sięga wypełnienie odpowiednim kolorem. Myląca może też być legenda. Kolorem ciemnoniebieskim powinniśmy mieć oznaczony przychód ogółem. Natomiast tutaj tym kolorem jest oznaczona konkretna część przychodów z wyłączeniem wpływów z reklam.

Poprawa wykresu

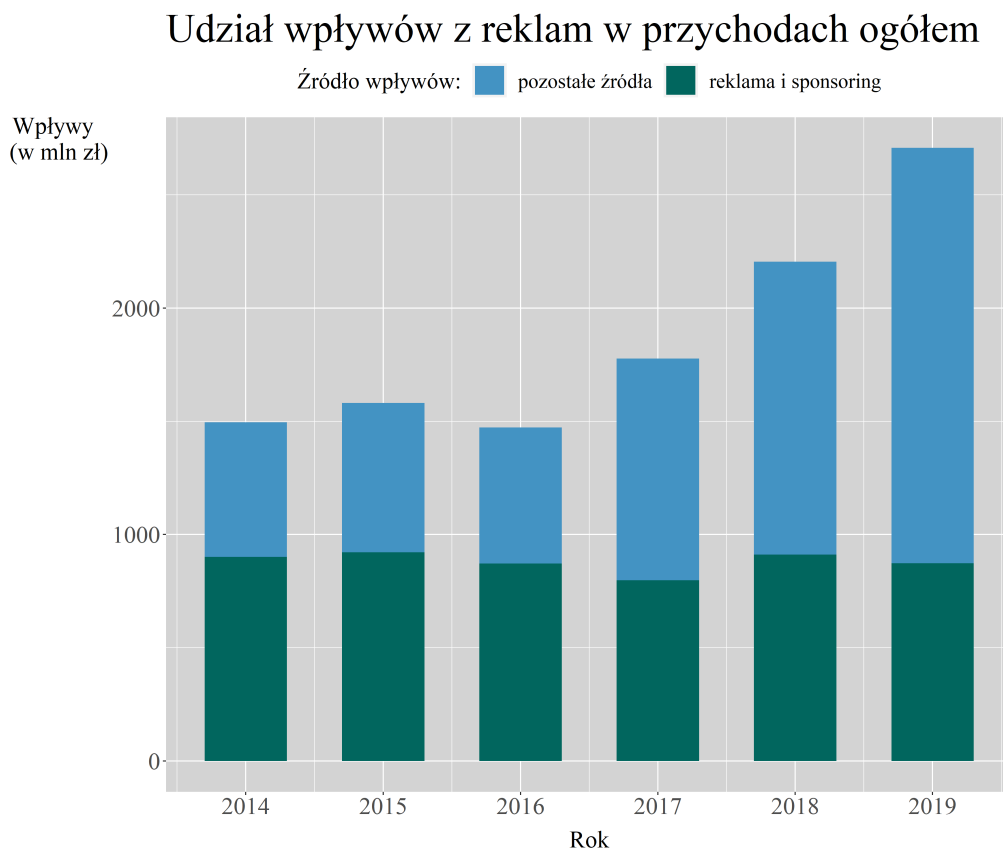
Kod w R

```
library(ggplot2)
library(dplyr)
library(extrafont)
loadfonts(device = "win")

ggplot(dane) +
  geom_col(aes(x = Rok, y = wartosc, fill = Zrodlo_wplywow), size = 10,
    width = 0.6) +
  scale_fill_manual(values = c("#4393c3", "#01665e")) +
  scale_x_continuous(breaks = seq(2014, 2019),
    labels = c("2014", "2015", "2016", "2017", "2018", "2019")) +
  labs(
    title = "Udzial_wplywow_z_reklam_w_przychodach_ogolem",
    fill = "Zrodlo_wplywow:",
    y = "Wplywy_n(w mln zl)",
    x = "Rok"
  ) +
  theme(
    panel.background = element_rect(fill = "lightgrey"),
    legend.title = element_text(size = 19),
    legend.text = element_text(size = 17, family = "serif"),
    legend.position = "top",
    axis.title.x = element_text(size = 20, margin = margin(t = 10)),
    axis.title.y = element_text(size = 20, angle = 0),
    axis.text.x = element_text(size = 20, family = "serif"),
    axis.text.y = element_text(size = 20, family = "serif"),
    title = element_text(size = 28, family = "serif"),
  )
ggsave("Corrected_plot.png", width = 30, height = 25, units = "cm")
```

Wykorzystane dane tabelaryczne pochodzą ze strony raportu, zostały najpierw zapisane w formacie csv, a następnie przetworzone tak, aby pozwoliły stworzyć wykres w ggplot2.

Wizualizacja



Rysunek 2: Wykres poprawiony

Uzasadnienie

Uważam, że użycie zwykłego kolumnowego wykresu do wizualizacji danych NIK-u jest korzystniejsze, ponieważ odbiorca znacznie szybciej poradzi sobie z ich odczytem. Wielkości wpływów są wyrażone wysokością słupka, dzięki czemu łatwiej jest je odczytać z osi. Natychmiastowo możemy też porównać, jak zmieniały się wpływy środków na przestrzeni lat, a także jaki procentowy udział w przychodzie miały reklamy. Zbędne stają się dodatkowe napisy na słupkach lub obok nich. Zmiana kategorii “przychód ogółem” na “pozostałe źródła” lepiej oddaje zawartość wykresu, jednocześnie nie zmieniając merytoryki. Od strony estetyki, wykres jest bardziej minimalistyczny, ale za to elegancki i informatywny, więc spełnia swoje najważniejsze zadanie.