

Praca domowa 3

Bartosz Szymański

1. Wizualizacja źródłowa



źródło: <https://twitter.com/TVP3Warszawa/status/1772648692773683425/photo/1>

Wizualizacja pochodzi z wpisu na platformie X przez TVP3 Warszawa z 26 marca 2024 roku.

(Na dzień dzisiejszy profil TVP usunął wpis z wykresem, jednak poniżej załączam link do jednego z wielu artykułów, potwierdzających, że rzeczywiście taki wpis miał miejsce. Powyżej dodatkowo screen, który sam wykonałem 27 marca przed usunięciem postu).

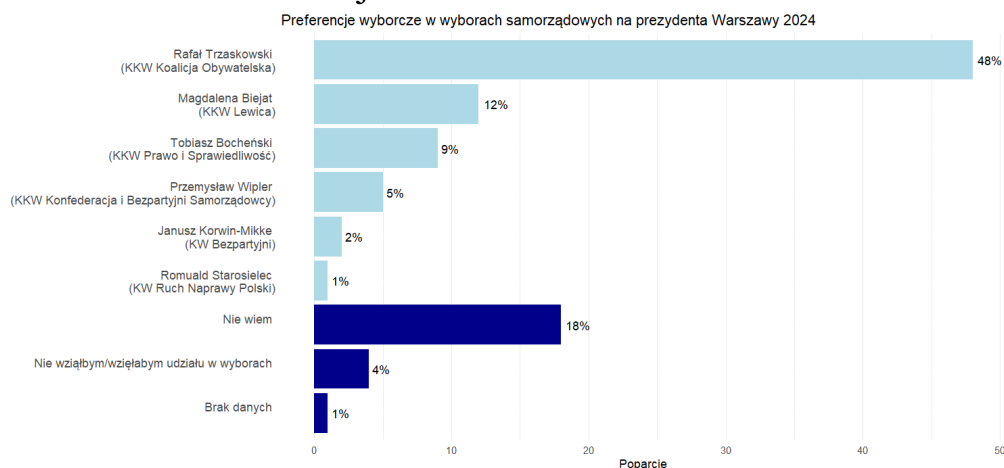
<https://kultura.gazeta.pl/kultura/7,127222,30836401,wykresowy-ze-starej-tvp-znow-atakuje-w-18-proc-az-trzy-razy.html?fbclid=IwAR3Z4RINCXpQKJVPuDCHFCdv2wze0JLqb854NkOcCX2uLt9PNKHn05VVR2I>

2. Elementy wymagające poprawy

Na wykresie zostały zupełnie zaburzone skale słupków (obrazek poniżej). Pierwszy z kandydatów ma dokładnie czterokrotnie większe poparcie niż druga kandydatka (licząc w punktach procentowych). Jednak słupek reprezentujący jego wynik jest znacząco większy niż czterokrotność poparcia drugiej kandydatki. Ponadto 18% reprezentujące niezdecydowanych wyborców jest niemal trzykrotnie większe niż 12% drugiej kandydatki. Taki wykres zaburza rzeczywiste postrzeganie wyników ankiety, sprawiając, że wynik części kandydatów wydaje się wyższy niż jest w rzeczywistości, a wynik pozostałych kandydatów wydaje się niższy. Dodatkowo suma punktów procentowych wszystkich opcji nie sumuje się do 100, a do 99 punktów procentowych. Ponadto kolor słupków zlewa się z tłem.



3. Utworzona wizualizacja



```
library(ggplot2)
library(dplyr)

data <- data.frame(
  kandydat = c("Brak danych", "Nie wziąłbym/wzięłabym udziału w wyborach", "Nie wiem", "Romuald Starosielec",
    "Janusz Korwin-Mikke", "Przemysław Wipler", "Tobiasz Bocheński", "Magdalena Biejał", "Rafał Trzaskowski"),
  Partia = c("", "", "", "(KW Ruch Naprawy Polski)", "(KW Bezpartyjni)", "(KKW Konfederacja i Bezpartyjni Samorządowcy)",
    "(KKW Prawo i Sprawiedliwość)", "(KKW Lewica)", "(KKW Koalicja Obywatelska)"),
  Poparcie = c(1, 4, 18, 1, 2, 5, 9, 12, 48)
)

data$kandydat <- factor(data$kandydat, levels = data$kandydat)

colors <- c(rep("darkblue", 3), rep("lightblue", 6))

data %>%
  ggplot(aes(x = Poparcie, y = kandydat)) +
  geom_col(fill = colors) +
  theme_minimal() +
  theme(
    panel.grid.major.y = element_line(color = "white"),
    axis.text.y = element_text(size = 11),
  ) +
  labs(title = "Preferencje wyborcze w wyborach samorządowych na prezydenta Warszawy 2024", y = "", x = "Poparcie") +
  geom_text(aes(y = kandydat, label = paste0(Poparcie, "%")), hjust = -0.2) +
  xlim(0, 55) +
  scale_y_discrete(labels = paste(data$kandydat, "\n", data$Partia)) +
  scale_x_continuous(breaks = seq(0, 60, by = 10))
```

4. Dlaczego nowy wykres jest lepszy?

Długości słupków w prawidłowych proporcjach oddają poparcie. Brakujący procent wynikał najprawdopodobniej z zaokrągleń wyników, jednak z racji braku możliwości weryfikacji danych (brak źródła do dokładnych danych), został przedstawiony w oddzielnym słupku jako brak danych, żeby nie faworyzować żadnego z kandydatów. Ponadto kolor słupków nie zlewa się już z otoczeniem.

Kod w wersji do skopiowania:

```
library(ggplot2)
library(dplyr)

data <- data.frame(
  Kandydat = c("Brak danych", "Nie wziąłbym/wzięłabym udziału w wyborach", "Nie wiem", "Romuald Starosielec",
    "Janusz Korwin-Mikke", "Przemysław Wipler", "Tobiasz Bocheński", "Magdalena Biejat", "Rafał Trzaskowski"),
  Partia = c("", "", "", "(KW Ruch Naprawy Polski)", "(KW Bezpartyjni)", "(KKW Konfederacja i Bezpartyjni Samorządowcy)",
    "(KKW Prawo i Sprawiedliwość)", "(KKW Lewica)", "(KKW Koalicja Obywatelska)"),
  Poparcie = c(1,4,18,1,2,5,9,12,48)
)

data$Kandydat <- factor(data$Kandydat, levels = data$Kandydat)

colors <- c(rep("darkblue", 3), rep("lightblue",6))

data %>%
  ggplot(aes(x = Poparcie, y = Kandydat)) +
  geom_col(fill = colors) +
  theme_minimal() +
  theme(
    panel.grid.major.y = element_line(color = "white"),
    axis.text.y = element_text(size = 11),
  ) +
  labs(title = "Preferencje wyborcze w wyborach samorządowych na prezydenta Warszawy 2024", y = "", x = "Poparcie")+
  geom_text(aes(y = Kandydat, label = paste0(Poparcie, "%")), hjust = -0.2) +
  xlim(0, 55) +
  scale_y_discrete(labels = paste(data$Kandydat, "\n", data$Partia)) +
  scale_x_continuous(breaks = seq(0, 60, by = 10))
```