

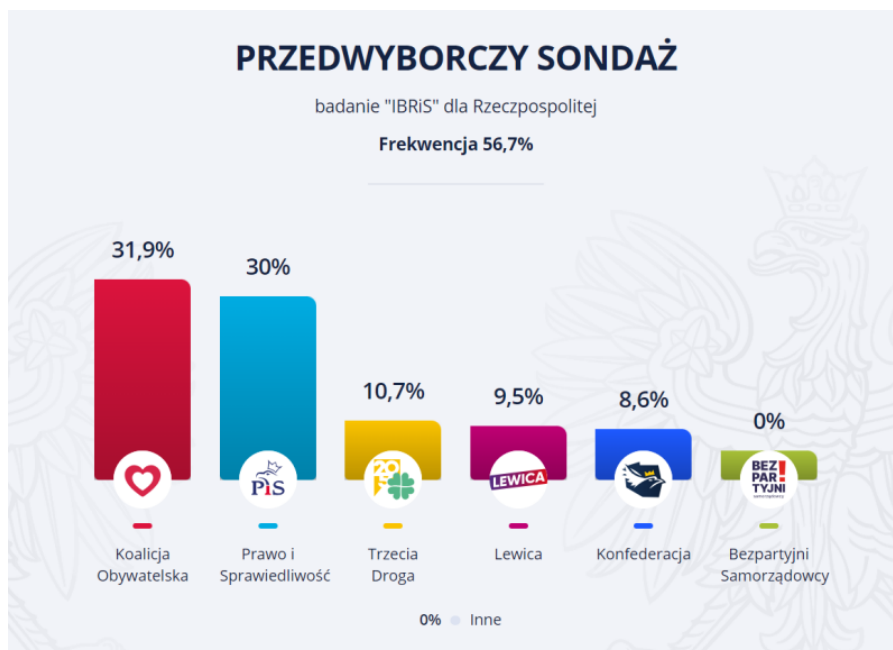
# Wstęp do Eksploracji Danych - HW 3

Marek Mączka

Kwiecień 2024

## 1 Wizualizacja źródłowa, która będzie poprawiana wraz z podaniem źródła i daty publikacji

Wykres, który zamierzam poprawić pochodzi ze strony pod tym linkiem. Został on opublikowany w środę, 27. marca 2024 r. o godzinie 07:25.



Rysunek 1: Znaleziony wykres

Badanie przeprowadzono metodą CATI na próbie 1067 osób w dniach 22-23 marca.

## 2 Krótkie uzasadnienie, jakie elementy wizualizacji wymagają poprawy

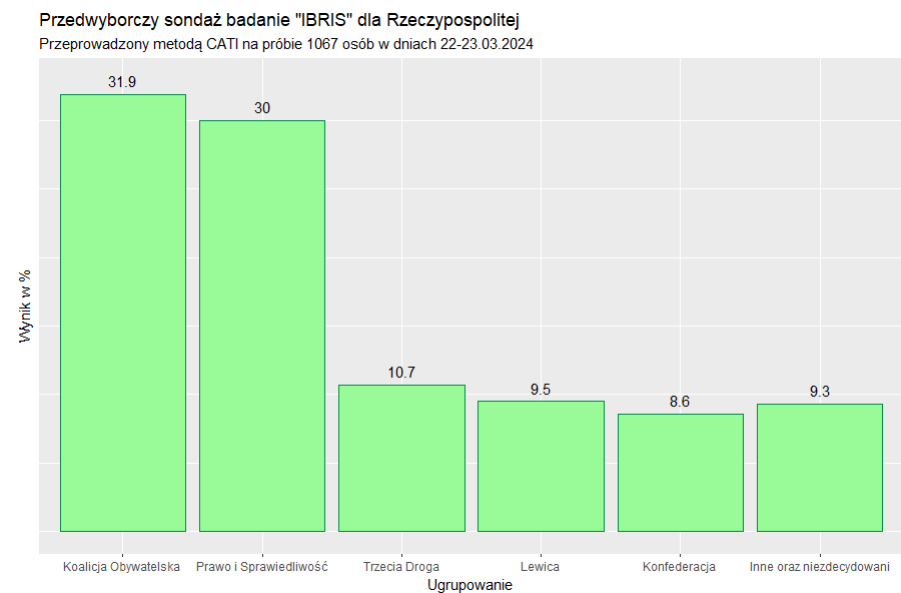
Przede wszystkim ciężko jest ocenić, gdzie jest punkt zerowy (początkowy) osi pionowej. Po dłuższym przyjrzeniu można dostrzec, że leży on na wysokości górnej części logo, a nie tam gdzie wszystkie prostokąty się zaczynają.

Druga sprawa to powtórzenie kilka razy informacji o partiach (lub koalicjach), poprzez ich nazwę, logo oraz wybrany kolor.

Kolejnym problemem jest niezdecydowanie w tym, czy wynik bliski 0% przedstawiać. Również sam napis 0% jest bardzo mylący.

Ostatnim problemem jest to, że wyniki nie sumują się do 100. Z danych z wykresu wychodzi 90,7%, co znacząco odbiega od pożądanej wartości. Ta nieścisłość nie została poruszona w raporcie na stronie. Trudno zatem stwierdzić, z czego to wynika. Jedynym moim podejrzeniem jest opcja wyboru w ankiecie odpowiedzi "Niezdecydowany/a".

## 3 Utworzona wizualizacja oraz kod potrzebny do jej wygenerowania



Rysunek 2: Utworzona wizualizacja

```

1 library(dplyr)
2 library(ggplot2)
3
4 data <- data.frame(Wynik = c(31.9, 30, 10.7, 9.5, 8.6, 9.3),
5                     Ugrupowanie = c("Koalicja Obywatelska", "Prawo i Sprawiedliwość",
6                                     "Trzecia Droga", "Lewica", "Konfederacja",
7                                     "Inne oraz niezdecydowani"))
8 data %>%
9   ggplot(aes(x=reorder(Ugrupowanie, -Wynik[c(1, 2, 3, 4, 6, 5)]), y=Wynik)) +
10   geom_col(fill="palegreen", color="springgreen4")+
11   labs(title = 'Przedwyborczy sondaż badanie "IBRIS" dla Rzeczypospolitej',
12        subtitle = 'Przeprowadzony metodą CATI na próbie 1067 osób w dniach 22-23.03.2024',
13        x = "Ugrupowanie",
14        y = "Wynik w %")+
15   geom_text(aes(y = Wynik+1, label = Wynik))+
16   theme(axis.text.y=element_blank(),
17         axis.ticks.y=element_blank())

```

Rysunek 3: kod potrzebny do wygenerowania wizualizacji

## 4 Krótkie uzasadnienie, dlaczego przygotowany wykres jest lepszy od oryginalnego

Rozwiązany został problem początku osi pionowej - teraz znajduje się w intuicyjnym miejscu. Informacje o partiach nie są powtórzone - zostawiłem aby nazwy. Z uwagi na brak informacji w raporcie pozostałe odpowiedzi poza pięcioma najczęściej wybieranymi zebrałem w jedną grupę i przedstawiłem na wykresie. Dzięki temu wynik w końcu sumuje się do 100%.

W ten sposób naprawiłem wszystkie ze znalezionych na początku błędów.