

Techniki Wizualizacji Danych

Praca domowa nr 2

Filip Langiewicz

Wizualizacja źródłowa:



Pochodzi ona z posta **SSPW - Samorząd Studentów Politechniki Warszawskiej** w wydarzeniu **Wybory do Organów Samorządowych na kadencję 2024** na Facebooku.

Źródło:

<https://www.facebook.com/photo/?fbid=718295917001060&set=gm.1037688767372152&id=1024933608647668>

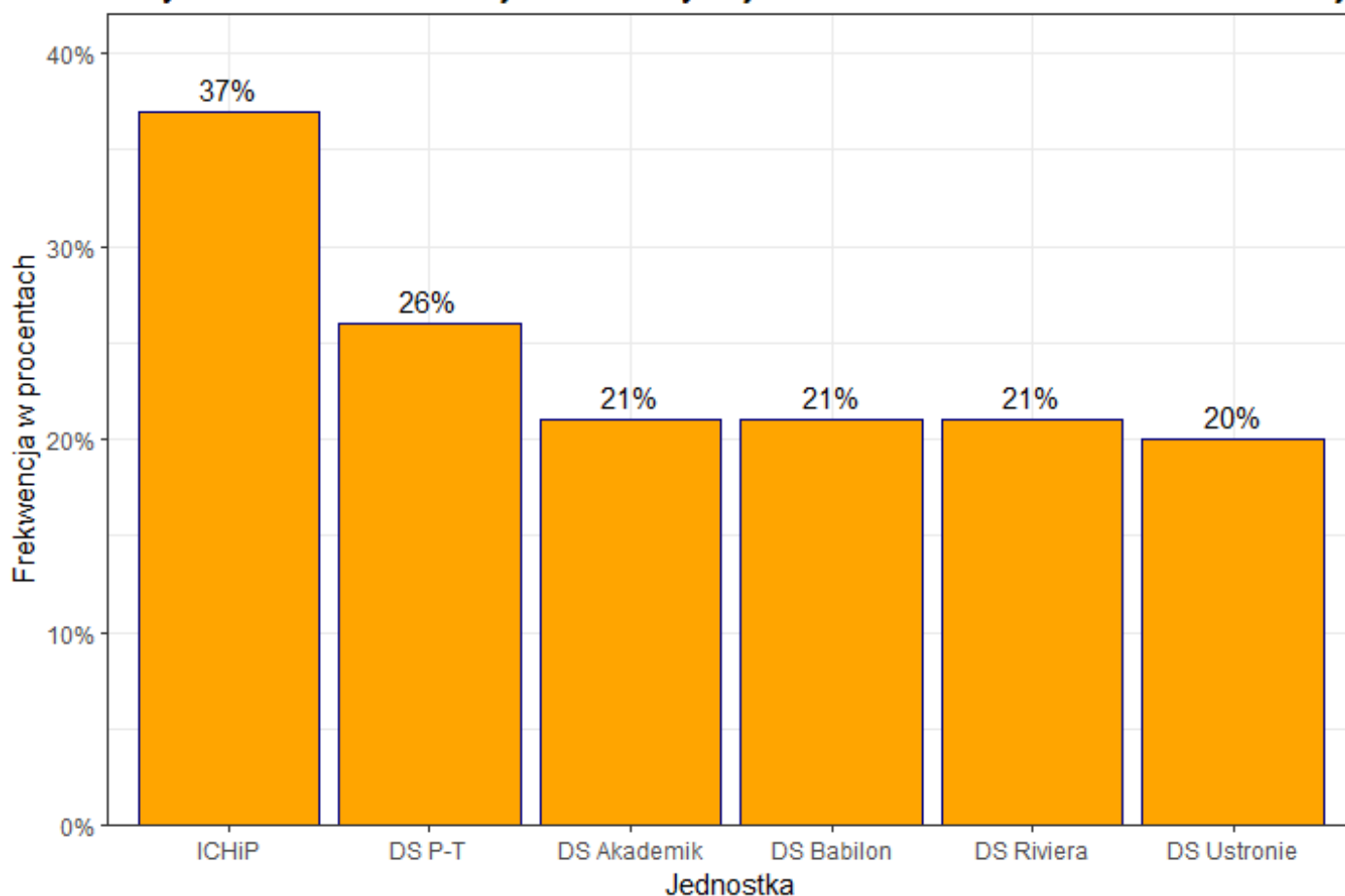
Data publikacji: 2 listopada 2023 roku.

Co wymaga poprawy?

Według mnie na powyższym wykresie poprawy wymaga wiele rzeczy. Między innymi nie są zachowane proporcje pomiędzy wysokościami poszczególnych słupków – słupek oddający frekwencję w DS P-T powinien być dłuższy od słupka DS Ustronie o $\frac{1}{4}$ jego długości, a już na pierwszy rzut oka tak nie jest, co sugeruje inne różnice między wartościami niż w rzeczywistości. Pomarańczowa dolna linia zdaje się imitować oś poziomą, od której ta wysokość powinna być mierzona, jednak słupki rozmywają się na dole i tak naprawdę nie wiadomo, w którym momencie się kończą, brakuje więc osi. Wyróżnia się również słupek najbardziej na lewo – nie jest opisany i tak naprawdę nie wiadomo co oznacza. Wychodzi on również poza ramkę całego wykresu, przez co jeszcze bardziej zaburza postrzeganie całości. Na koniec należy wspomnieć, że trzy słupki oddają tę samą wartość – 21%, lecz nie mają tej samej wysokości, co potęguje brak rzetelności w przedstawieniu danych.

Poprawiona wizualizacja (bez słupka najbardziej na lewo – nie wiadomo co oznacza):

Dotychczasowa frekwencja w niektórych jednostkach Politechniki Warszawskiej



Kod potrzebny do jej uzyskania:

```
1 library(ggplot2)
2 library(forcats)
3 library(scales)
4
5 df <- data.frame("Jednostka" = c("ICHiP",
6                                "DS P-T",
7                                "DS Akademik",
8                                "DS Babilon",
9                                "DS Riviera",
10                               "DS Ustronie"),
11                 "Procent" = c(37,
12                               26,
13                               21,
14                               21,
15                               21,
16                               20))
17
18 ggplot(df, aes(x = fct_reorder(Jednostka, -Procent), y = Procent)) +
19   geom_col(fill = "orange",
20           color = "navy") +
21   geom_text(aes(label = paste(Procent, "%", sep = "")),
22           vjust = -0.5) +
23
24   labs(title = "Dotychczasowa frekwencja w niektórych jednostkach Politechniki Warszawskiej",
25        x = "Jednostka",
26        y = "Frekwencja w procentach") +
27   scale_y_continuous(expand = expansion(c(0, 0), c(0, 5)),
28                     labels = percent_format(scale = 1)) +
29   theme_bw()
30
31
32
```

Dlaczego jest lepsza niż oryginalna?

Poprawiony wykres jest lepszy od oryginalnego, ponieważ przedstawia dane w sposób schludny i łatwy do odczytu. Ponadto zachowane są tu proporcje dzięki czemu oglądający może prawidłowo odczytać zależności pomiędzy poszczególnymi frekwencjami, np. widać od razu zdecydowaną różnicę między frekwencją w DS Ustronie a ICHiP – em, co nie było oczywiste na oryginalnym wykresie.