

Praca domowa 2 TWD

Techniki wizualizacji danych

Samsel Aleksandra

2023-11-03

Wizualizacja i źródła

Poniższa wizualizacja pochodzi z gazety “Przegląd sportowy WEEKEND”, numer 88 (27-29.10.2023 r.)



Co jest nie tak?

Wizualizacja porównuje ze sobą frekwencję średnią na jeden mecz oraz sumaryczną dla kolejnych sezonów rozgrywek. Mylące dla mnie jest to, że słupki są bardzo podobnych wysokości pomimo, że słupki czerwone przedstawiają wartości rzędu tysięcy a zielone - milionów. Na pierwszy rzut oka ciężko zrozumieć dlaczego mniejsze wartości są przedstawione przez wyższe słupki.

Poprawiony wykres

```
library(ggplot2)
library(dplyr)
library(ggpubr)

df <- data.frame (
  MeanFreq = c(8878, 9030, 4949, 7373, 9405, 11937),
  SumFreq = c(2610002, 2037884, 248489, 2249435, 2877795, 1253333),
  Year = c('2018/19', '2019/20', '2020/21', '2021/22', '2022/23', '2023/24'),
  NumberOfRounds = c(37, 37, 30, 34, 34, 12)
)
```

```

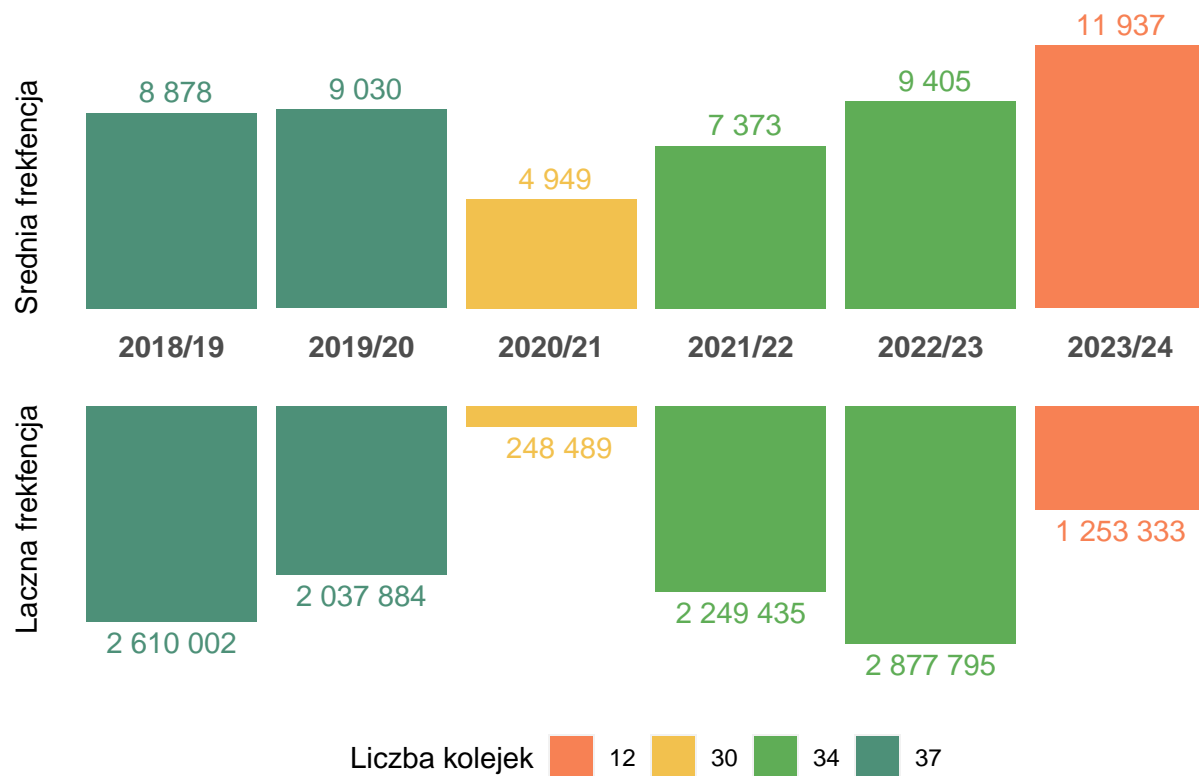
gMean <- df %>%
  ggplot(aes(x = Year, y = MeanFreq, fill = as.factor(NumberOfRounds))) +
  geom_col() +
  ggtitle("FREKFENCJA W EKSTRAKLASIE") +
  ylab("Średnia frekfencja") +
  ylim(0, 13000) +
  theme(legend.position = "none",
        plot.title = element_text(face = "bold", color = '#4a4a4a'),
        axis.ticks.y = element_blank(),
        axis.title.y = element_text(hjust = .1),
        axis.text.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x = element_blank(),
        axis.text.x = element_text(size = 11, face = 'bold'),
        panel.background = element_rect(fill = 'white',
                                          color = 'white')) +
  geom_text(aes(label = format(MeanFreq, big.mark = " ",
                              scientific = FALSE),
                color = as.factor(NumberOfRounds)), vjust = -.5,
            size = 4) +
  scale_fill_manual(values =
                    c("#f78154", "#f2c14e", "#5fad56", "#4d9078")) +
  scale_colour_manual(values =
                      c("#f78154", "#f2c14e", "#5fad56", "#4d9078"))

gSum <- df %>%
  ggplot(aes(x = Year, y = SumFreq, fill = as.factor(NumberOfRounds))) +
  geom_col() +
  ylab("Łączna frekfencja") +
  scale_y_reverse(limits = c(3200000,0)) +
  theme(legend.position="bottom",
        axis.ticks.y = element_blank(),
        axis.title.y = element_text(hjust = .8),
        axis.text.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        panel.background = element_rect(fill = 'white',
                                          color = 'white')) +
  geom_text(aes(label = format(SumFreq, big.mark = " ",
                              scientific = FALSE),
                color = as.factor(NumberOfRounds)), vjust = 1.5, size=4,
            show.legend = FALSE) +
  guides(fill = guide_legend(title = "Liczba kolejek")) +
  scale_fill_manual(values =
                    c("#f78154", "#f2c14e", "#5fad56", "#4d9078")) +
  scale_colour_manual(values =
                      c("#f78154", "#f2c14e", "#5fad56", "#4d9078"))

ggarrange(gMean, gSum, nrow = 2, ncol = 1)

```

FREKFENCJA W EKSTRAKLASIE



Co zostało poprawione?

Uważam, że poprawiony wykres jest mniej mylący, ponieważ łączna i średnia frekwencja zostały widocznie rozłączone na dwa wykresy. Dodatkowo informację o liczbie kolejek w danym roku można przedstawić za pomocą kolorów, co wpływa na walory estetyczne wykresu.