

Praca domowa 2

Urszula Szczęsna

2023-11-07

Znaleziony wykres do poprawienia

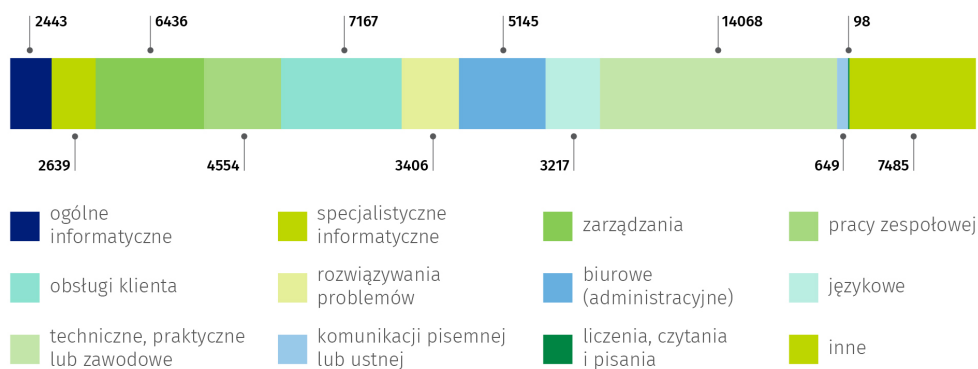
Wykres który znalazłam został udostępniony przez GUS na twitterze w dniu 20 października 2023.

https://twitter.com/GUS_STAT/status/1715308324101447751/photo/1

(https://twitter.com/GUS_STAT/status/1715308324101447751/photo/1).



Liczba przedsiębiorstw, których pracownicy w 2020 r. nabyli umiejętności na kursach:



@GUS_STAT

@GłównyUrządStatystyczny

Główny Urząd Statystyczny

@gus_stat

Główny Urząd Statystyczny GUS

Błędy w wykresie źródłowym

Kolory na wykresie są bardzo zbliżone, co znacząco utrudnia szybkie odczytanie dokładnych wartości.

Ponadto, umiejętności nie są uporządkowane, co sprawia, że wybranie na przykład trzech umiejętności, które zostały nabyte przez największą ilość pracowników, stanowi wyzwanie i zajmuje sporo czasu. Dodatkowo, porównywanie ze sobą dwóch umiejętności jest trudne, ponieważ nie są one przedstawione na wspólnej osi.

Poprawianie wykresu

Kod potrzebny do stworzenia odpowiedniej ramki danych

```
# załadowanie potrzebnych bibliotek
library(dplyr)
```

```
## Warning: pakiet 'dplyr' został zbudowany w wersji R 4.2.3
```

```
library(ggplot2)
```

```
## Warning: pakiet 'ggplot2' został zbudowany w wersji R 4.2.3
```

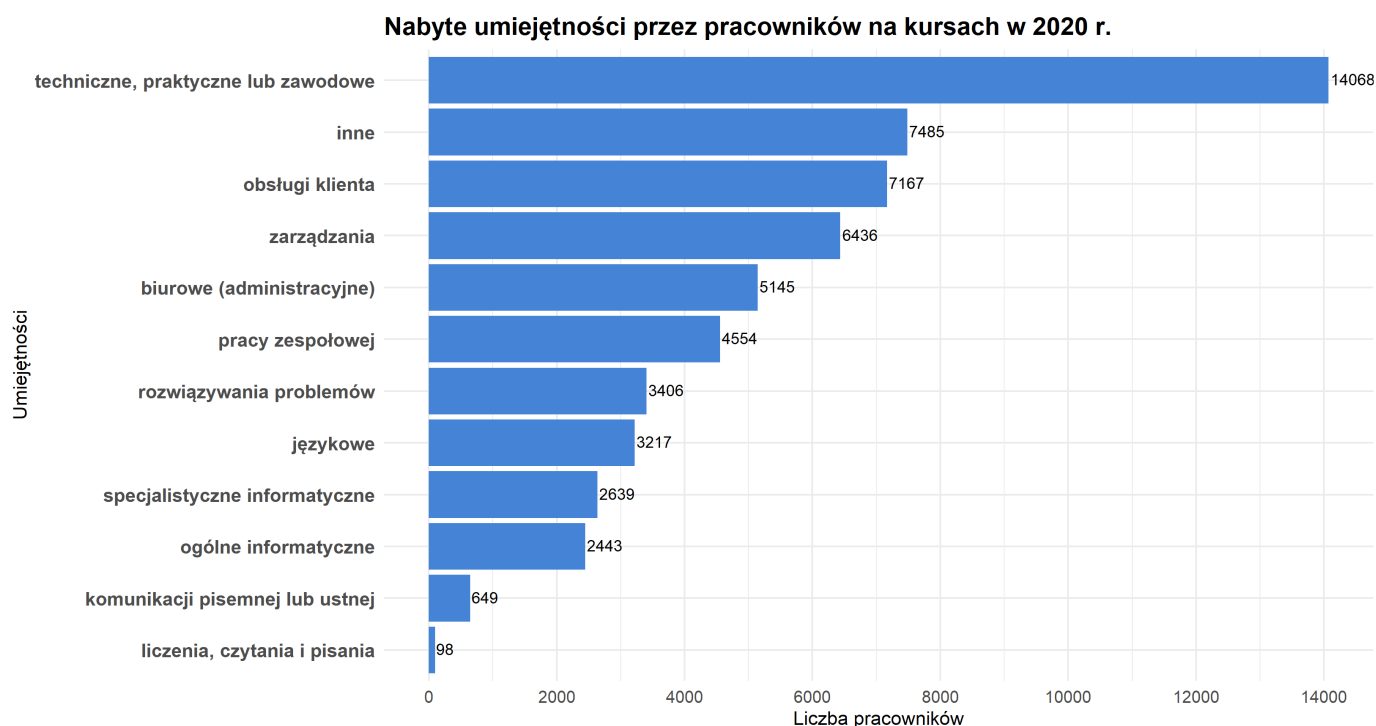
```
# tworzenie data frame, które będzie potrzebne do wygenerowania wykresu
df <- data.frame(y = c("ogólne informatyczne", "specjalistyczne informatyczne",
                      "zarządzania", "pracy zespołowej",
                      "obsługi klienta", "rozwiązywania problemów",
                      "biurowe (administracyjne)", "językowe",
                      "techniczne, praktyczne lub zawodowe",
                      "komunikacji pisemnej lub ustnej",
                      "liczenia, czytania i pisanie", "inne"),
                 x = c(2443, 2639, 6436, 4554, 7167, 3406, 5145, 3217, 14068,
                      649, 98, 7485 ))
colnames(df) <- c("skill", "number")

# posortujemy skill malejąco po number
df %>%
  arrange(-number) -> df
```

Kod potrzebny do wygenerowania wykresu

```
ggplot(df, aes(y = reorder(skill, number), x = number)) +
  geom_col(fill = "#4483d5" )+
  geom_text(aes(label = number), vjust = 0.4, hjust = -0.05, size = 5) +
  labs( title = "Nabyte umiejętności przez pracowników na kursach w 2020 r. ",
        x = "Liczba pracowników",
        y = "Umiejętności"
  )+
  scale_x_continuous(breaks = seq(0, max(df$number), by = 2000)) +
  theme_minimal(base_size = 17) +
  theme(plot.title = element_text(size = 22, face = "bold"),
        axis.text.y = element_text(size = 17, face = "bold"),
        axis.text.x = element_text(size = 15),
  ) -> plot

plot
```



Uzasadnienie

Nowy wykres jest znacznie lepszy, ponieważ jest to wykres słupkowy, który doskonale podkreśla, które umiejętności zdobyło najwięcej pracowników. Dzięki wykresowi słupkowemu w prosty i szybki sposób możemy porównać dane liczbowe. Dodatkowo, umieszczenie dokładnych wartości na szczycie każdego słupka umożliwia natychmiastowe odczytanie konkretnych liczb. Wszystkie słupki mają ten sam kolor, ze względu na obecność opisów na osi Y, co eliminuje konieczność kolorowania słupków. Dzięki temu wykres jest znacznie bardziej czytelny i przejrzysty.