

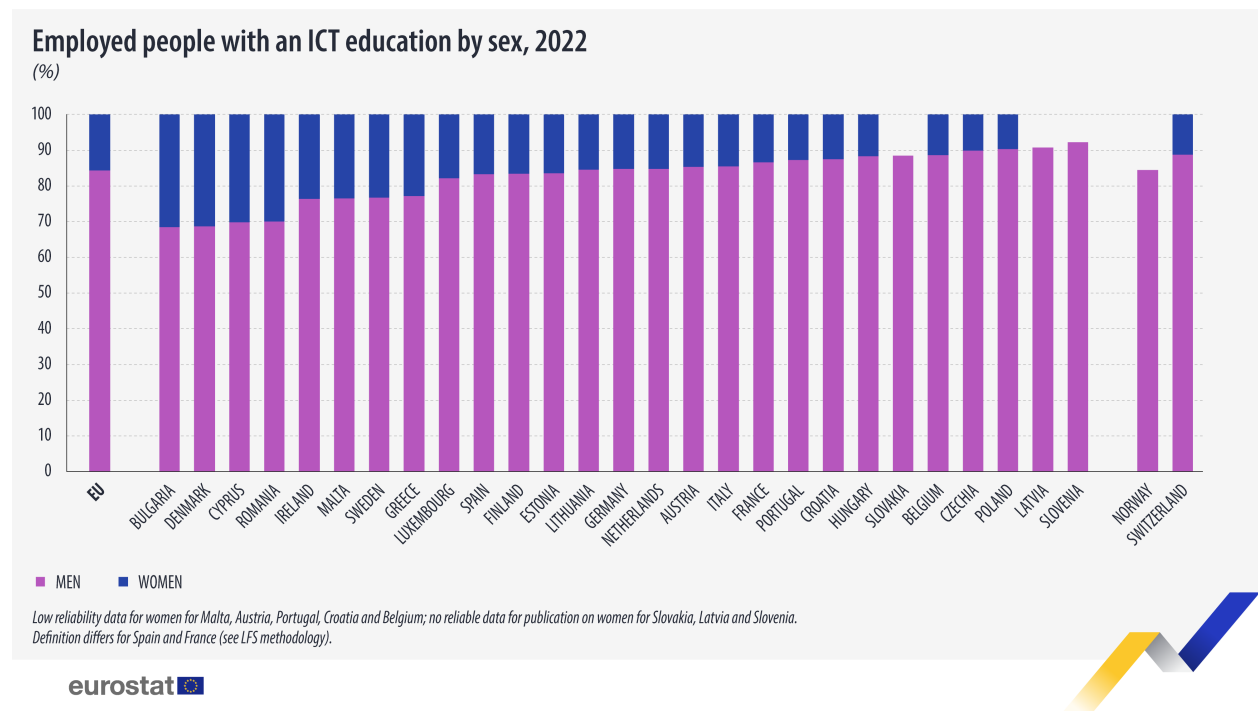
Techniki wizualizacji danych. Zadanie 4.

Gaspar Sekula

11.12.2023

1 Wizualizacja oryginalna

Oryginalna wizualizacja (**Rysunek 1.**) pochodzi z artykułu *Men represented 84% of people employed with an ICT education* opublikowanego na stronie Eurostatu dnia 16. października 2023. Tekst i wykres dostępne są tutaj.



Rysunek 1: Oryginalny wykres.

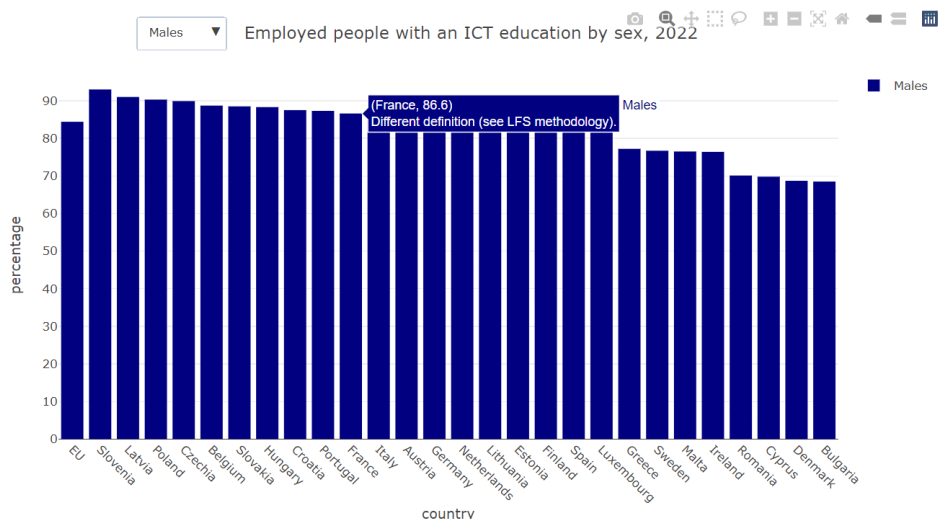
2 Co wymaga poprawy?

Można powiedzieć, że poprawianie błędów wizualizacji przeprowadzonych przez Eurostat to donkiszoteria, bowiem są one w większości bardzo dobre i precyzyjne. Lecz zaprezentowany powyżej wykres wymaga korekty, mianowicie:

- Dla prawie 30. wartości na osi poziomej wykres staje się nieczytelny, ponieważ kolumny są zbyt wąskie, jest ich zbyt dużo i są zbyt blisko siebie.

- Linie podziałki skali osi pionowej są niemalże niewidoczne (szczególnie pomiędzy kolumnami), co uniemożliwia odczytanie dokładnej wartości.
- Jeśli mamy dane dla mężczyzn, łatwo obliczyć wartość dla kobiet. Można więc wykres sprowadzić do wykresu procentowego udziału kobiet (albo mężczyzn) z wykształceniem IT wśród zatrudnionych.

3 Propozycja poprawy



Rysunek 2: Poprawiony wykres.

Powyższy wykres powstał na podstawie danych podanych jako źródło wykresu oryginalnego (**Rysunek 1.**). Poprawiona wizualizacja (**Rysunek 2.**) realizuje następujące zmiany na lepsze:

- ✓ Tradycyjny wykres kolumnowy jest teraz interaktywny - możliwe jest łatwe odczytanie interesującej nas wartości poprzez najechanie kursorem na odpowiedni słupek.
- ✓ Można wybrać płeć, dla której dane chcemy odczytywać. Możliwe jest również wybranie obu płci jednocześnie, lecz taki wykres nie ma sensu (dla każdego państwa suma wartości dla kobiet i mężczyzn sumuje się do 100).
- ✓ Informacje ze stopki wykresu o niedokładnych danych lub odmiennej metodologii zostały przeniesione do opisu słupków. Nie zwracają zbędnie naszej uwagi, a są dostępne tam, gdzie będą przydatne.
- ✓ Poprawiony wykres prezentuje dane dla jednej płci (zależnie jaką opcję wybierzemy), lecz przy założeniach, że są dwie płcie (co zostało przyjęte przy tworzeniu oryginalnego wykresu) łatwo odczytać proporcje dla tej drugiej (lub zmienić ustawienia wyświetlania).
- ✓ Przy skrupulatnym wyszukiwaniu interesujących nas danych w oryginalnej wizualizacji, może dochodzić do zmęczenia oczu z powodu wykorzystania stosunkowo agresywnych barw. W poprawionej wersji zaproponowano łagodniejsze kolory.
- ✓ Podobnie jak w oryginalnej wizualizacji, średnia wartość dla państw członkowskich UE jest wyszczególniona jako pierwszy element.

4 Kod generujący wykres

Kod w R potrzebny do wygenerowania wykresu z **Rysunku 2.** znajduje się w katalogu `my_plot`.

```
1 ### TWD - HW 4 ###
2
3 ### libraries
4
5 library(dplyr)
6 library(plotly)
7
8 ### data frame
9
10 df <- read.csv("data.csv")
11
12 ### modify data frame
13
14 df$OBS_VALUE <- as.numeric(df$OBS_VALUE)
15
16 df <- df %>%
17   select("sex", "geo", "OBS_VALUE", "OBS_FLAG") %>%
18   mutate( geo = ifelse(geo == "European Union - 27 countries (from 2020)", "EU",
19     geo),
20     OBS_FLAG = case_when(OBS_FLAG == "d" ~ "Different definition (see LFS
21       methodology).",
22       OBS_FLAG == "u" ~ "Low reliability data for women
23       in this country.",
24       TRUE ~ ""))
25
26 df_ <- df
27
28 df <- df %>% arrange(desc(OBS_VALUE)) %>%
29   filter(!is.na(OBS_VALUE) & geo != "EU")
30
31 df_ <- df_ %>% arrange(desc(OBS_VALUE)) %>%
32   filter(!is.na(OBS_VALUE) & geo == "EU")
33
34 df <- rbind(df_, df)
35
36 df$geo <- factor(df$geo, levels = rev(unique(df$geo)))
37
38 ### plot
39
40 plot_ly(data = df,
41   x = ~geo,
42   y = ~OBS_VALUE,
43   type = "bar",
44   color = ~sex,
45   legendgroup = ~sex,
46   text = ~OBS_FLAG,
47   colors = c("darkred", "navy")) %>%
48   layout(
49     title = "Employed people with an ICT education by sex, 2022",
50     xaxis = list(
51       title = "country",
52       tickangle = 45,
53       categoryorder = "total",
54       categoryarray = df$geo
55     ),
```

```

53 yaxis = list(title = "percentage"),
54
55 updatemenus = updatemenus <- list(
56   list(
57     buttons = list(
58       list(method = "restyle",
59         args = list("visible", list(TRUE, TRUE)),
60         label = "Both"),
61       list(method = "restyle",
62         args = list("visible", list(FALSE, TRUE)),
63         label = "Males"),
64       list(method = "restyle",
65         args = list("visible", list(TRUE, FALSE)),
66         label = "Females")
67     ),
68     direction = "down",
69     showactive = TRUE,
70     x = 0.1,
71     xanchor = "left",
72     y = 1.15,
73     yanchor = "top"
74   )
75 )
76 )
77 )

```

Listing 1: Kod generujący wykres