

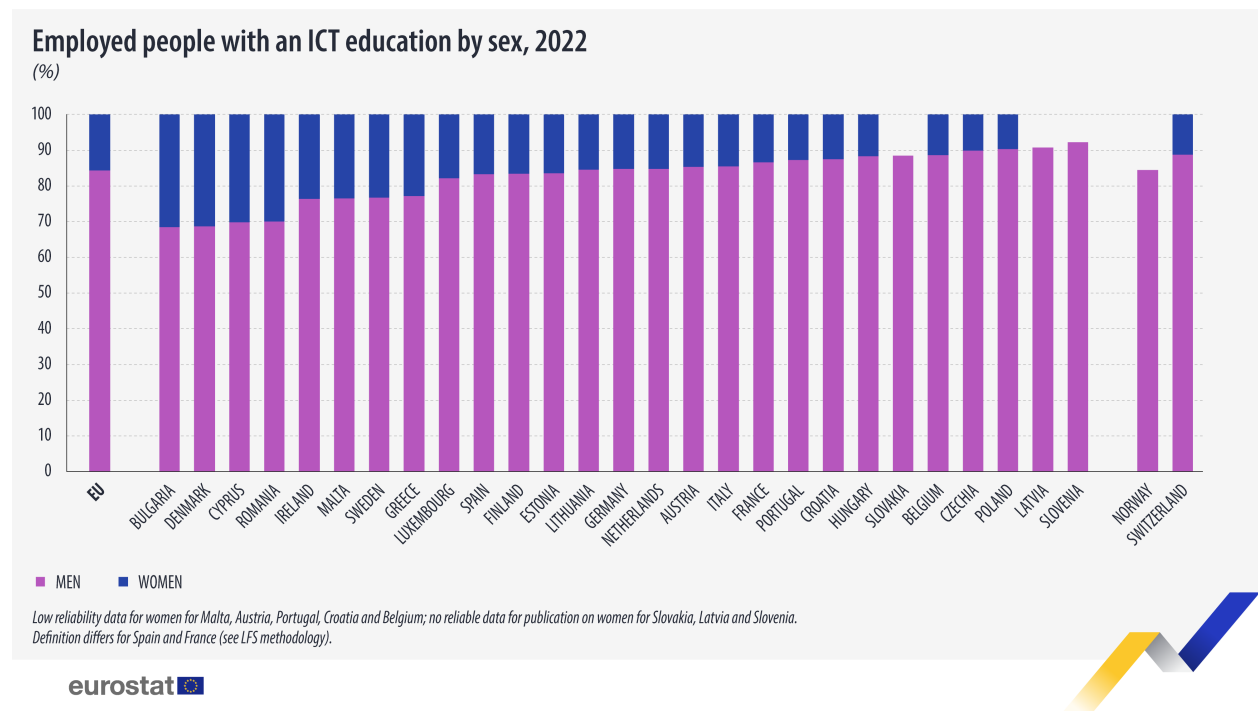
Techniki wizualizacji danych. Zadanie 2.

Gaspar Sekula

November 4, 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:

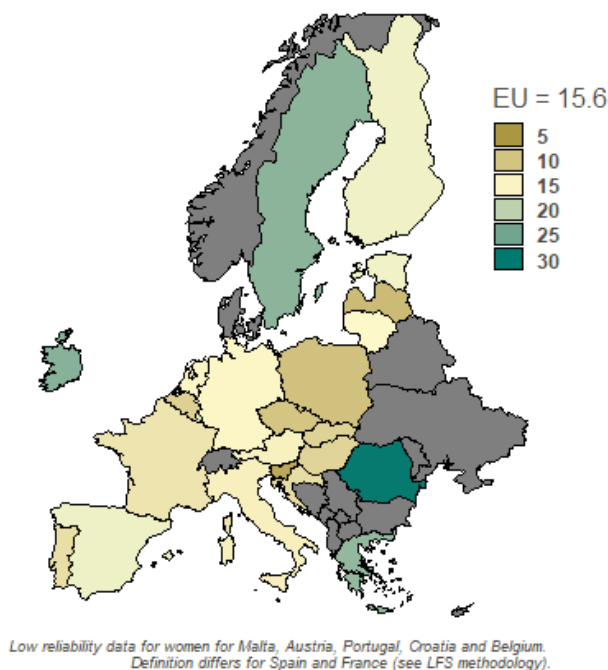
- Dla prawie 30. wartości na osi poziomej wykres staje się nieczytelny, ponieważ kolumny są zbyt wąskie, jest ich zbyt dużo i są zbyt blisko siebie.

- ☛ Linie podziałki skali osi pionowej są niemalże niewidoczne (szczególnie pomiędzy kolumnami), co uniemożliwia odczytanie dokładnej wartości.
- ☛ Dla prawie 30. państw bardzo trudno jest szybko znaleźć dane dotyczące interesującego nas kraju.
- ☛ Informacja o roku badania (2022) ma istotnie mniejsze znaczenie niż pozostała część tytułu, toteż może znaleźć się w podtytule.

3 Propozycja poprawy

Employed women with ICT education

as percentage of all employed people with ICT education, 2022



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:

- ✓ Zastąpienie wykresu kolumnowego mapą zwiększa jego czytelność - nie ma lasu 30. słupków blisko siebie. Co więcej, łatwo można badać zależności zatrudnienia kobiet (lub mężczyzn) z wykształceniem informatycznym od położenia państwa, itd.
- ✓ Ułatwione porównywanie państw między sobą. Dobrana skala kolorów umożliwia sprawne zestawienie wyników dla wybranych krajów.
- ✓ W źródłowej ramce danych nie ma wartości dla Norwegii i Szwajcarii.

- ✔ Poprawiony wykres prezentuje dane dla kobiet, lecz przy założeniach, że są dwie płcie (co zostało przyjęte przy tworzeniu oryginalnego wykresu) łatwo odczytać proporcje dla mężczyzn.
- ✔ 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.
- ✔ Mniej istotne informacje (w szczególności rok) w tytule zostały przeniesione do podtytułu.

4 Kod generujący wykres

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

```

1 ### libraries
2
3 library(dplyr)
4 library(tidyr)
5 library(ggplot2)
6 library(maps)
7 library(viridis)
8
9 df <- read.csv("data.csv")
10
11 ### modify data frame
12
13 prepared <- df %>% filter((sex == "Males") & (!is.na(OBS_VALUE))) %>%
14   select(geo, OBS_VALUE, OBS_FLAG) %>%
15   mutate(Females = 100 - OBS_VALUE) %>%
16   mutate(geo = ifelse(geo == "Czechia", "Czech Republic", geo))
17
18 ### prepare data frame
19
20 european_countries <- c(
21   "Albania", "Andorra", "Austria", "Belarus", "Belgium", "Bosnia and Herzegovina",
22   "Bulgaria", "Croatia", "Cyprus", "Czech Republic", "Denmark", "Estonia", "
23     Finland",
24   "France", "Germany", "Greece", "Hungary", "Ireland", "Italy", "Kosovo",
25   "Latvia", "Liechtenstein", "Lithuania", "Luxembourg", "Malta", "Moldova", "
26     Monaco",
27   "Montenegro", "Netherlands", "North Macedonia", "Norway", "Poland", "Portugal",
28   "Romania", "San Marino", "Serbia", "Slovakia", "Slovenia", "Spain", "Sweden",
29   "Switzerland", "Ukraine", "United Kingdom", "Vatican City"
30 )
31
32 eu_map <- map_data("world", region = european_countries) %>%
33   filter(lat < 70)
34
35 eu_lab_data <- eu_map %>%
36   group_by(region) %>%
37   summarise(longitude = mean(long), latitude = mean(lat))
38
39 plot_data <- left_join(eu_map, prepared, by = join_by(region == geo))
40
41 ### generate plot
42

```

```

43 ggplot(plot_data, aes(x = long, y = lat)) +
44   geom_polygon(aes( group = group, fill = Females), color = "black") +
45   theme_void()+
46   scale_fill_gradient2(
47     low = "gold4", high = "#01796f", mid = "lemonchiffon",
48     limits = c(5, 30), midpoint = 15.6,
49     guide = guide_legend( keyheight = unit(4, units = "mm"),
50                           keywidth=unit(5, units = "mm"),
51                           label.position = "right",
52                           title.position = 'top',
53                           nrow=6,
54                           title = "EU = 15.6"
55                           )
56   )+
57   labs(title = "Employed women with ICT education",
58        subtitle = "as percentage of all employed people with ICT education, 2022",
59        caption = "Low reliability data for women for Malta, Austria, Portugal,
60                  Croatia and Belgium.
61                  Definition differs for Spain and France (see LFS methodology).
62                  ") +
63   theme(
64     text = element_text(color = "black"),
65     plot.background = element_rect(fill = "white", color = NA),
66     panel.background = element_rect(fill = "white", color = NA),
67     legend.background = element_rect(fill = "white", color = NA),
68     plot.title = element_text(size=16, hjust = 0.01, color="black", face = "bold
69     ",
70     margin = margin(b = -0.1, t = 0.4, l = 2, unit = "
71     cm")),
72     plot.subtitle = element_text(size= 10, hjust=0.01, color = "#4e4d47",
73     margin = margin(b = -0.1, t = 0.43, l = 2, unit
74     = "cm")),
75     plot.caption = element_text( size=7, color = "#4e4d47", margin =
76     margin(b = 0.3, r=-99, unit = "cm"), face = "
77     italic"),
78     legend.position = c(1, 0.68),
79     legend.title = element_text(color = "#4e4d47", size = 12),
80     legend.text = element_text(color = "#4e4d47", face = "bold", size = 9)
81   ) +
82   coord_map()

```

Listing 1: Kod generujący wykres