

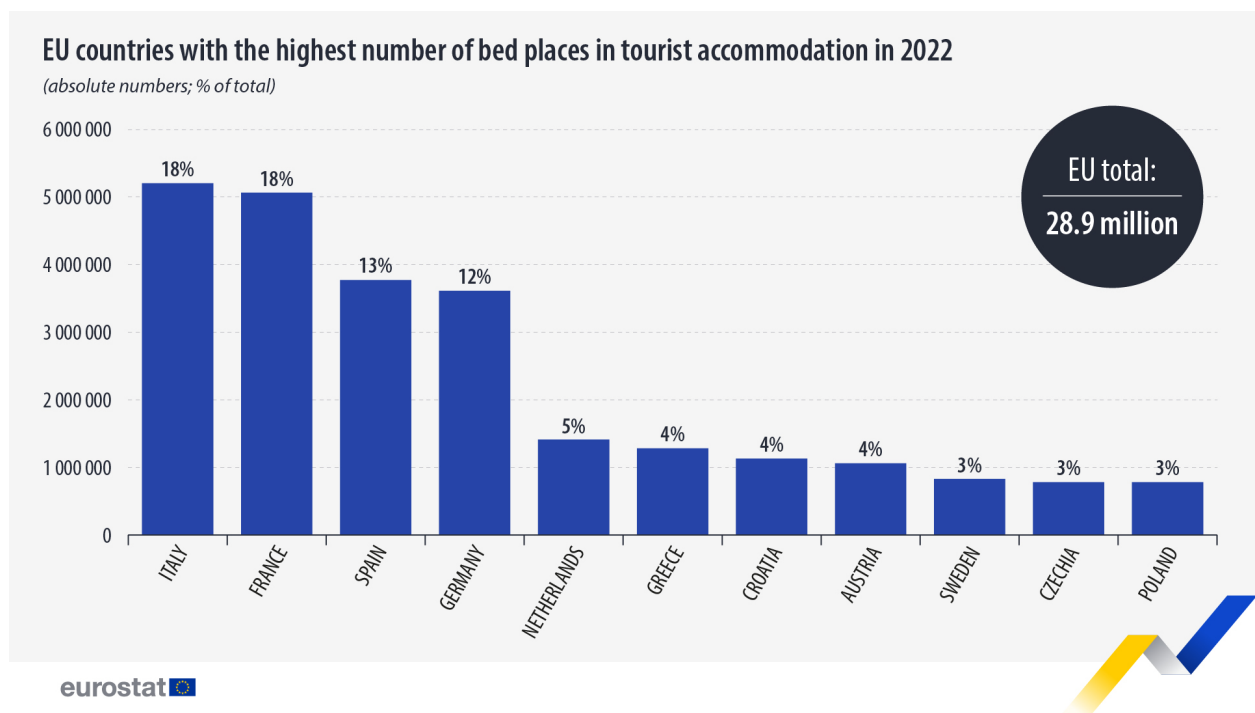
# Techniki wizualizacji danych - Praca domowa nr 2

Filip Szlingiert

November 4, 2023

## 1 Wizualizacja źródłowa

Wizualizacja (**Rysunek 1.**) pochodzi z artykułu *EU sees increase of tourist bed places in 2022* opublikowanego na oficjalnej stronie Eurostatu 31 października 2023. Artykuł, wykres oraz link do danych znajdują się tutaj.



Rysunek 1: Oryginalny wykres.

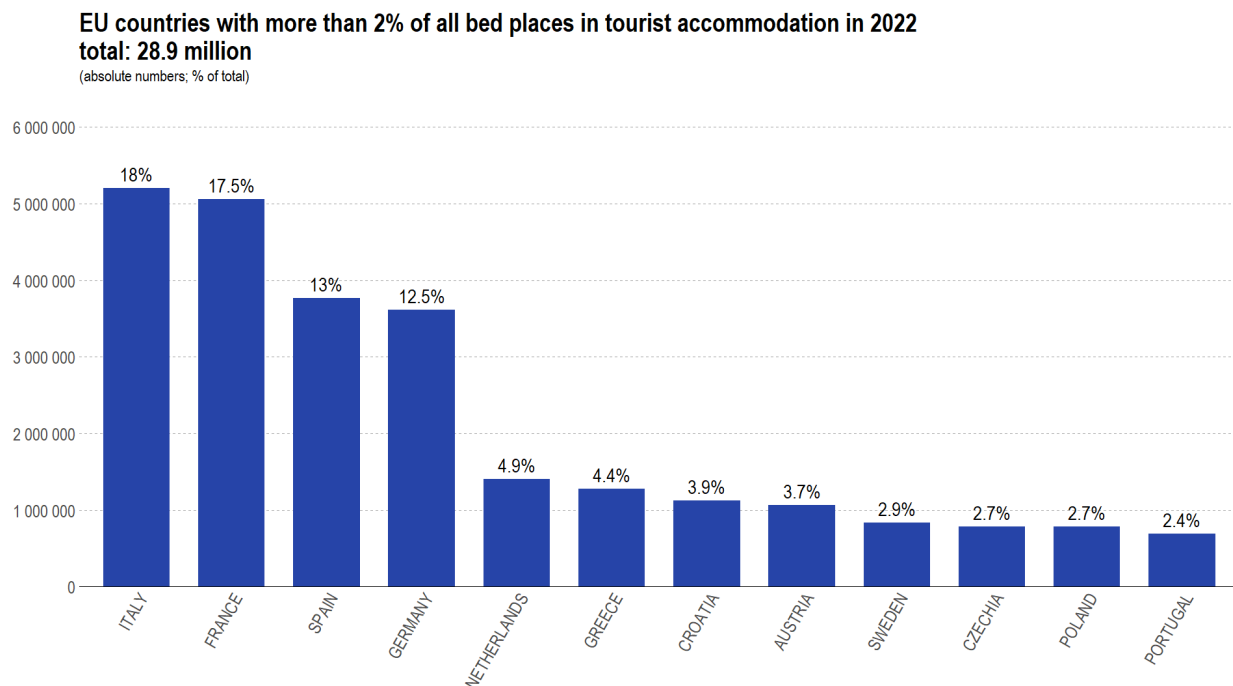
## 2 Co można poprawić?

Na pierwszy rzut oka wykres wydaje się nie mieć żadnych błędów, typowy wykres słupkowy, natomiast przy ponownych spojrzeniu można zauważyć kilka niedociągnięć:

- ✗ Słupki o różnych wysokościach mają te same wartości procentowe co poddaje pod wątpliwość poprawność danych.

- ✗ Autorzy nie sprecyzowali co oznacza zwrot "państwa z największą liczbą łóżek". Czytelnik może pomyśleć, że wykres przedstawia państwa od tego o najlepszym wyniku w UE do tego o najgorszym, przez co błędnie oceni Polskę lub Czechy. Może też się zastanawiać jak wygląda sytuacja pozostałych państw i jakie kryterium zadecydowało o przedstawieniu właśnie tych państw, gdyż równie dobrze wykres mógłby się skończyć na Szwecji.

### 3 Moja propozycja poprawy.



Rysunek 2: Moja wizualizacja.

Wykres powstał na podstawie danych w artykule oznaczonych jako źródło oryginalnego wykresu (**Rysunek 1.** ). Moja wizualizacja (**Rysunek 2.**) ma następujące zalety:

- ✓ Wprowadzenie wartości procentowych zaokrąglonych to jednego miejsca po przecinku sprawia, że różnice wysokości słupków o tych samych wartościach stają się niezauważalne.
- ✓ Tytuł w jasny sposób tłumaczy dlaczego inne państwa nie pojawiły się na wykresie. Zdecydowałem się dodać Portugalie, gdyż patrząc jak rozkładają się wartości, zauważyłem, że nie różni się ona znacznie od wartości dla Polski, natomiast po Portugalii następuje dosyć duży spadek, bo do 1.6 (spadek o 1/3).
- ✓ Została wykowana w pobnym stylu co wykres źródłowy.

### 4 Kod generujący wykres

1 `### Biblioteki`

2

```

3 library(tidyr)
4 library(dplyr)
5 library(ggplot2)
6 library(forcats)
7 library(extrafont)
8
9 font_import()
10 loadfonts(device = "win")
11
12
13 ### Wczytanie i praca z danymi
14
15 stats <- read.csv("tour_cap_nat__custom_8097699_page_linear.csv")
16
17
18 stats %>% select(geo, OBS_VALUE) %>%
19   arrange(-OBS_VALUE) %>%
20   filter(!is.na(OBS_VALUE)) -> Bedplaces_2022_tmp
21
22
23 ### Brak danych dla Irlandii - suma wszystkich lozek zostala oszacowana
24
25 Bedplaces_2022_tmp %>% top_n(1) -> tmp
26
27
28 estimated_sum <- tmp$OBS_VALUE
29
30
31 Bedplaces_2022_tmp %>%
32   filter(OBS_VALUE != estimated_sum) %>%
33   mutate(percent = OBS_VALUE / estimated_sum * 100) -> Bedplaces_2022
34
35
36 ### Sprawdzam, czy wstawienie oszacowanej wartosci dla Irlandii cos zmieni w mojej
   ramce
37 ### Okazuje sie, ze Irlandia znajdowalaby sie na jednym z ostatnich miejsc, wiec
   ja pomijam
38
39 Bedplaces_2022 %>%
40   summarise(all = sum(OBS_VALUE)) -> tmp2
41
42
43 sum_no_ireland <- tmp2$all
44
45
46 estimated_ireland <- estimated_sum - sum_no_ireland
47
48
49 ### Przystosowuje ramke
50
51 Bedplaces_2022 %>%
52   mutate(percent = round(percent, digits = 1),
53           geo = toupper(geo),
54           geo = fct_reorder(geo, -OBS_VALUE)) %>%
55   filter(percent > 2) %>%
56   mutate(percent = paste0(as.character(percent), "%"))-> for_plot
57
58
59 ### Tworze wykres

```

```

60
61 ggplot(for_plot, aes(x = geo,
62                       y = OBS_VALUE)) +
63   geom_col(width = 0.7, fill = "#2644a8") +
64   scale_x_discrete(guide = guide_axis(title = "",
65                                       angle = 60)) +
66   scale_y_continuous(breaks = c(0, 1000000, 2000000, 3000000, 4000000, 5000000,
67                                6000000),
68                     labels = c(0,
69                                format(1000000, big.mark = " ", scientific = FALSE
70                                ),
71                                format(2000000, big.mark = " ", scientific = FALSE
72                                ),
73                                format(3000000, big.mark = " ", scientific = FALSE
74                                ),
75                                format(4000000, big.mark = " ", scientific = FALSE
76                                ),
77                                format(5000000, big.mark = " ", scientific = FALSE
78                                ),
79                                format(6000000, big.mark = " ", scientific = FALSE
80                                )),
81   limits = c(0, 6500000),
82   expand = c(0,0),
83   guide = guide_axis(title = "")) +
84   theme_minimal() +
85   theme(text = element_text(family = "Arial Narrow"),
86         axis.text = element_text(size = 20),
87         plot.title = element_text(size = 30,
88                                   face = "bold"),
89         plot.subtitle = element_text(size = 18),
90         panel.grid.minor = element_blank(),
91         panel.grid.major.x = element_blank(),
92         panel.grid.major.y = element_line(linetype = "dashed",
93                                           colour = "darkgrey"),
94         axis.line.x = element_line()) +
95   labs(title = "EU countries with more than 2% of all bed places in tourist
96          accommodation in 2022\ntotal: 28.9 million",
97        subtitle = "(absolute numbers; % of total)") +
98   geom_text(aes(label = percent),
99             position = position_dodge(width = 0.5),
100            vjust = - 0.5,
101            family = "Arial Narrow",
102            size = 8)

```

Listing 1: Kod generujący wykres