

Wstęp

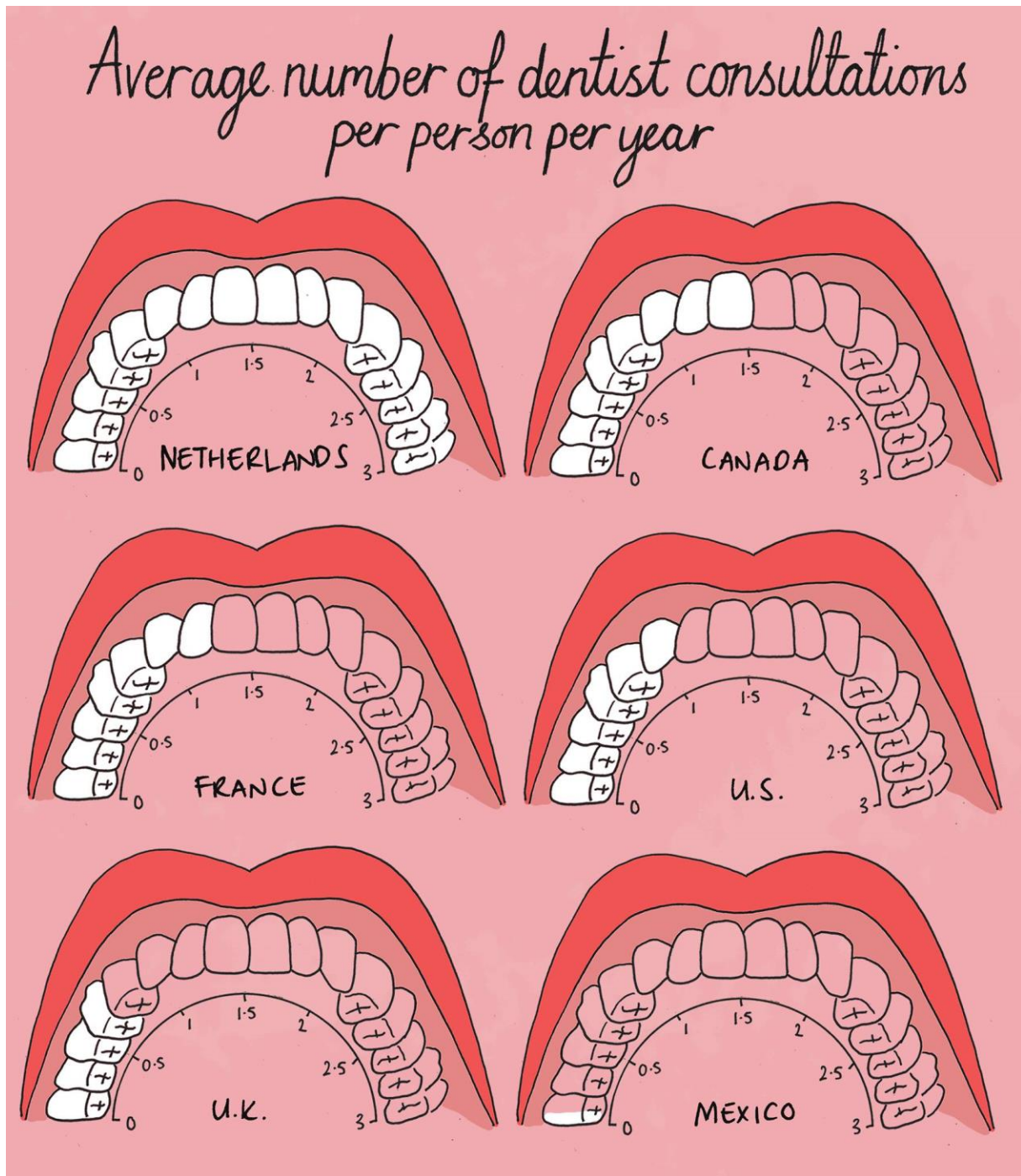
W niniejszym raporcie zamierzam omówić jeden wykres oraz wytłumaczyć, dlaczego (według mnie) nie jest on najbardziej czytelnym sposobem zaprezentowania danych. Autorem wizualizacji jest Mona Chalabi, został on zamieszczony 25. października br. przez stronę edukacyjną TED na Facebooku i przedstawia średnią liczbę wizyt mieszkańców 6 krajów u dentysty w ciągu roku.

W Internecie przytoczony wykres budzi kontrowersje – niektórzy uważają go za świetny, zwracając szczególną uwagę na walory estetyczne, kreatywność autorki oraz przystępność w odbiorze.

Mimo to, można znaleźć go na Reddicie na forum poświęconemu szczególnie złym wizualizacjom danych ([r/dataisugly](https://www.reddit.com/r/dataisugly)). Poniżej znajduje się moja analiza tego co w tym wykresie może się nie podobać oraz próbę innej, czytelniejszej (czy lepszej, to już zależy od gustu) wizualizacji.

Niestety nie wiadomo skąd pochodzą dane. Mam jednak bardzo mocne podstawy sądzić, iż z bazy OECD, gdyż wszystkie kraje zamieszczone na wykresie są członkami tej organizacji, a po *wygooglowaniu*, znalazłem dataset, który zawiera bardzo podobne wyniki (ciężko stwierdzić czy te same, bo skala wykresu uniemożliwia dokładne porównanie).

Ale o co chodzi?

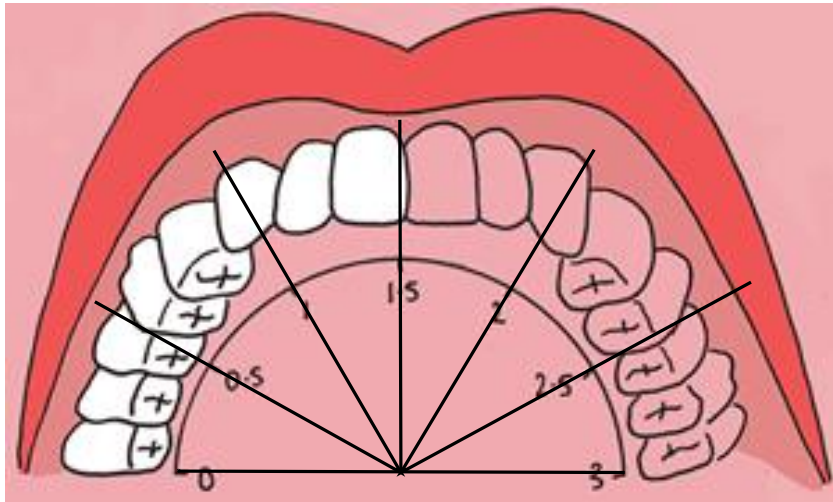


Omawiany wykres; Autor: Mona Chalabi

Teraz, gdy każdy miał już czas zapoznać się z wykresem, opiszę problemy jakie w nim widzę:

1. Na jeden ząb przypada mniej więcej (!) 3/16 wizyty, co może wprowadzić w błąd kogoś, kto nie ma czasu przyjrzeć się szczegółom osi i może potraktować jeden ząb np. za jedną wizytę.
2. Ząb nie jest najmniejszą jednostką miary (przykład Meksyku), co jeszcze bardziej utrudnia odbiór.
3. Ząb zębowi nie równy.
4. Ze względu na powyższe ciężko porównywać przedstawione dane oraz określić dokładną wartość liczbową.
5. Dane nie pochodzą z tego samego okresu, co nie zostało poruszone na wykresie.
6. Brak źródła danych.
7. Arbitralne wybranie krajów do wykresu, mimo dostępnych danych dla innych krajów.

W celu zilustrowania punktów 1-4 niech posłuży poniższy obrazek.

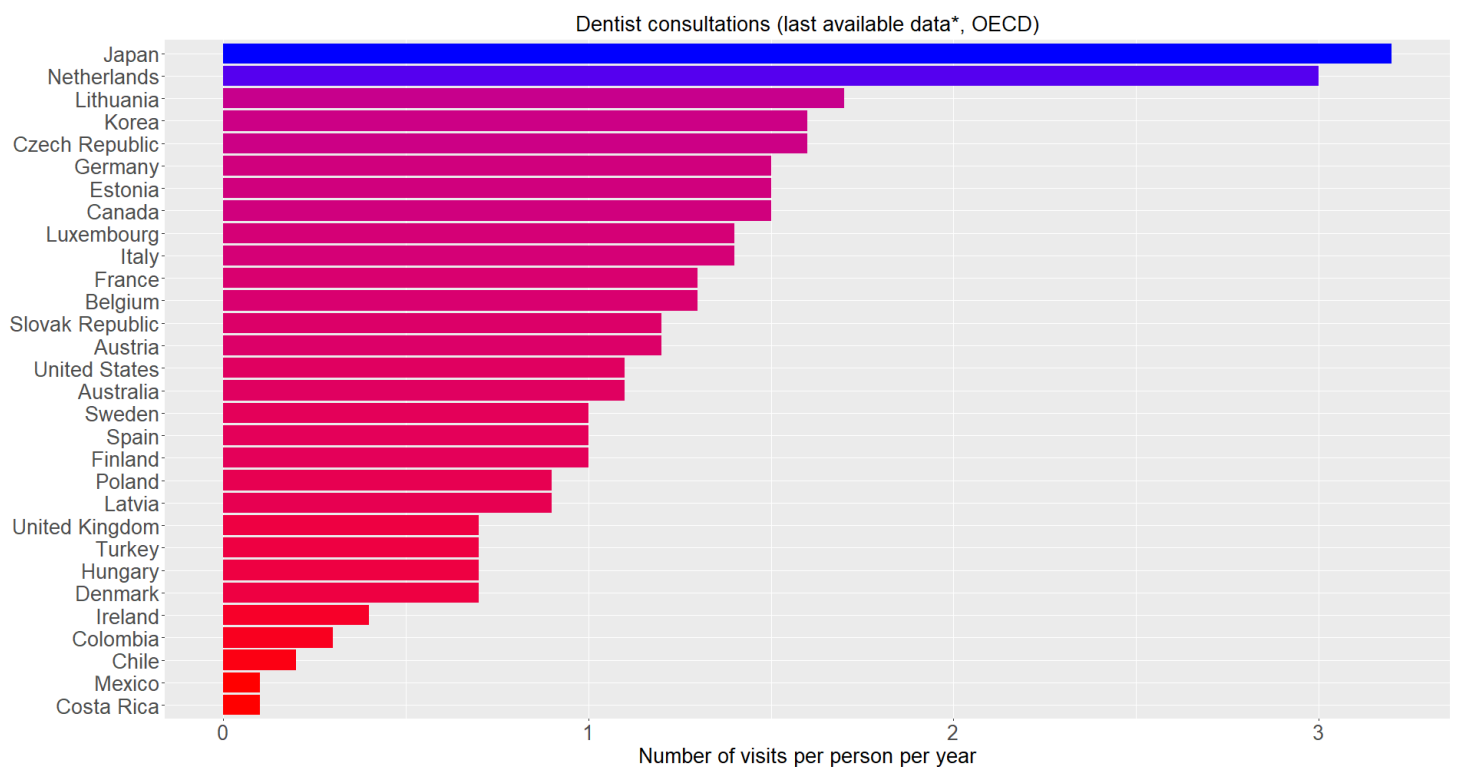


Dodanie linii pomocniczych; opracowanie własne

Jasno widać, że pierwszy i ostatni przedział skali zawiera trzy zęby, podczas gdy pozostałe przedziały zawierają 2.5 zęba.

Poprawiony wykres

Ostatnią część stanowi moja wizualizacja danych. Poszeregowałem kraje OECD, według liczby konsultacji z dentystą *per capita* dla ostatniego dostępnego okresu (nie uwzględniłem krajów, które ostatnie dane miały sprzed 2018). Dane umieściłem na wykresie słupkowym. Poniżej rezultat. Myślę, że mój wykres jest lepszy, ponieważ jest lepiej opisany, nie posiada mylącej skali oraz opowiada pełniejszą i mniej arbitralną historię.



Źródło: OECD; opracowanie własne

Kod

HW2 TWD

28.10.2021

Andrzej Pioro

importing libraries

library("dplyr")

library("ggplot2")

setting working directory and importing data

setwd("C:/Users/pioro/Documents/PW sem. 3/TWD/Homeworks/HW2")

consultations <- read.csv("Dentist_consultation_pc_OECD.csv")

wybór potrzebnych danych

consultations_latest <- consultations %>%

select(Year, Country, Value) %>%

group_by(Country) %>%

filter(Year == max(Year)) %>%

filter(Year > 2017)

tworzenie wykresu

p <- ggplot(data = consultations_latest) +

geom_bar(aes(

x = reorder(Country, Value),

y = Value,

fill = Value

```

),
stat = "identity",
show.legend = F) +
ggtitle("Dentist consultations (last available data*, OECD)") +
theme(plot.title = element_text(hjust = 0.5)) +
theme(
  axis.text = element_text(size = 20),
  axis.title = element_text(size = 20),
  plot.title = element_text(size = 20)
) +
xlab("") + ylab("Number of visits per person per year") +
coord_flip() +
scale_fill_gradient(low = "red", high = "blue") +
labs(caption = "*Countries with last data entries before 2018 were omitted") +
theme(plot.caption = element_text(size = 13))

```

wyświetlenie wykresu

p