

Kubacki_Jakub_313494_pd2

Jakub Kubacki

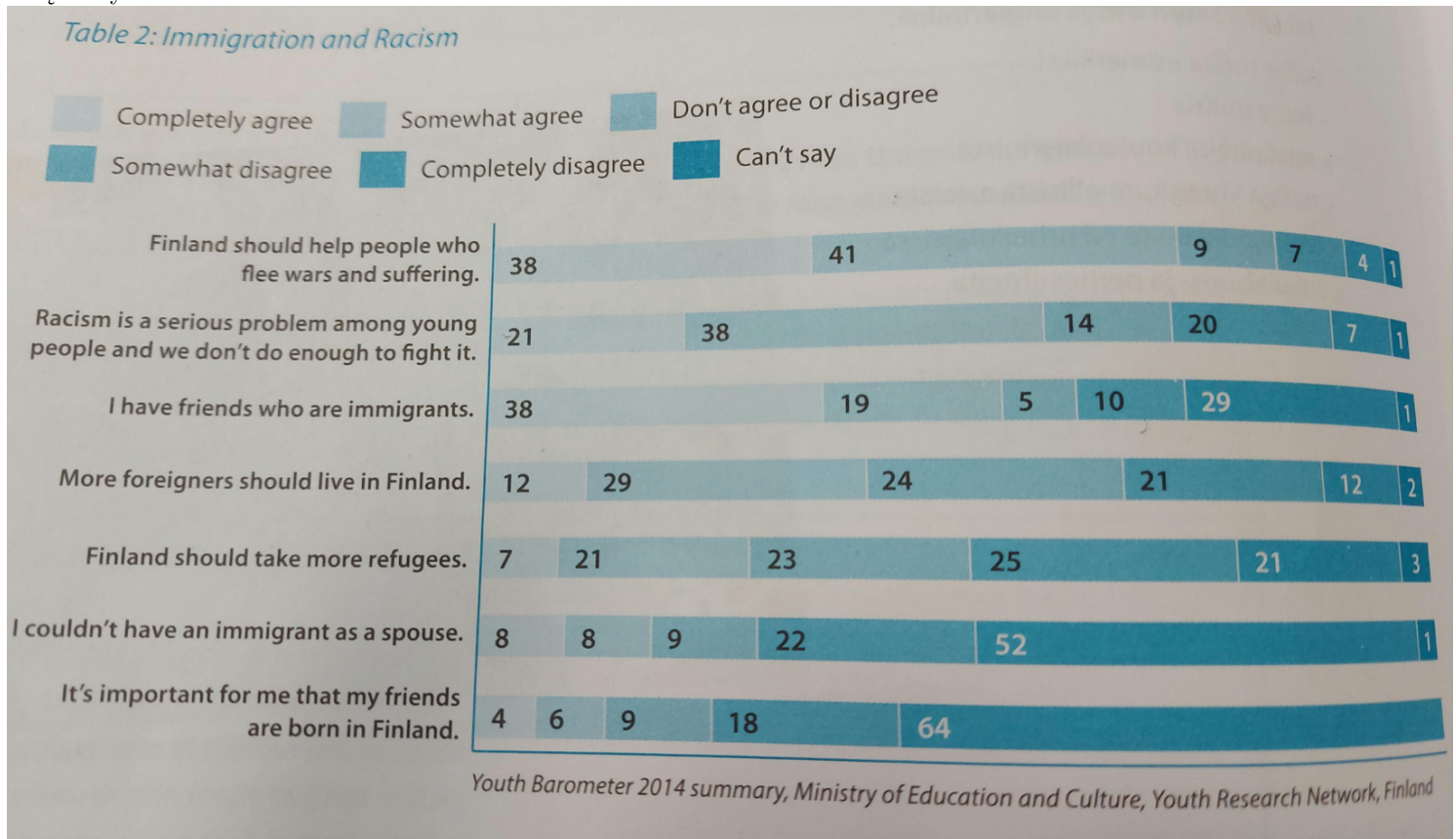
28 03 2022

Oryginalny wykres

Wykres, który znalazłem obrazuje (albo przynajmniej próbuje to robić) opinie Finów z zakresu imigracji i pokazuje skalę uprzedzeń wynikających z rasizmu i ksenofobii... w 2014 roku.

Wiem, że nie spełnia to wszystkich kryteriów, jakie zostały narzucone w pracy domowej, ale na moją obronę, pojawił się on na subreddicie "dataisugly" tydzień temu a wcześniej, w przeciwieństwie do wykresów z TVP, nie widziałem go nigdzie. Nie umiałem też znaleźć niczego, co spełniałoby wszystkie kryteria, więc zdecydowałem, że popracuję nad tym...

...A jest nad czym pracować. Nie będę wymieniał wszystkiego, co jest złe w tym wykresie. Po prostu go pokażę i oddam ocenę w ręce czytelnika:



Oczywiście w dalszej części raportu pokażę moją wersję wykresu i opowiem o wprowadzonych zmianach i dlaczego te zmiany uważałem za stosowne.

Tworzenie wykresu

Nie dostałem się do danych z samego badania, od którego powstał oryginalny wykres, ale jako, że wykres ma charakter dyskretny i mimo licznych problemów, jest podpisany, zdołałem przepisać wartości z wykresu do csv-ki.

W .csv tabela ma kształt 6-kolumnowej tabeli, gdzie każdy wiersz dotyczy wyników jednego pytania. Pierwsza kolumna mówi o treści pytania (gdzie pozwoliłem sobie zaprzeczyć treść dwóch z nich i zamienić stosowne wyniki, co uzasadnię później) a następne 5 zawierają przepisane z wykresu procenty odpowiedzi: “zdecydowanie się zgadzam”, “raczej się zgadzam”, “nie wiem” (suma “nie powiem” i “ani się nie zgadzam ani się nie-nie zgadzam”), “raczej się nie zgadzam” i “zdecydowanie się nie zgadzam”.

Korzystając z danych z tabelki i z poniższego ciągu poleceń wygenerowałem moją wersję wykresu:

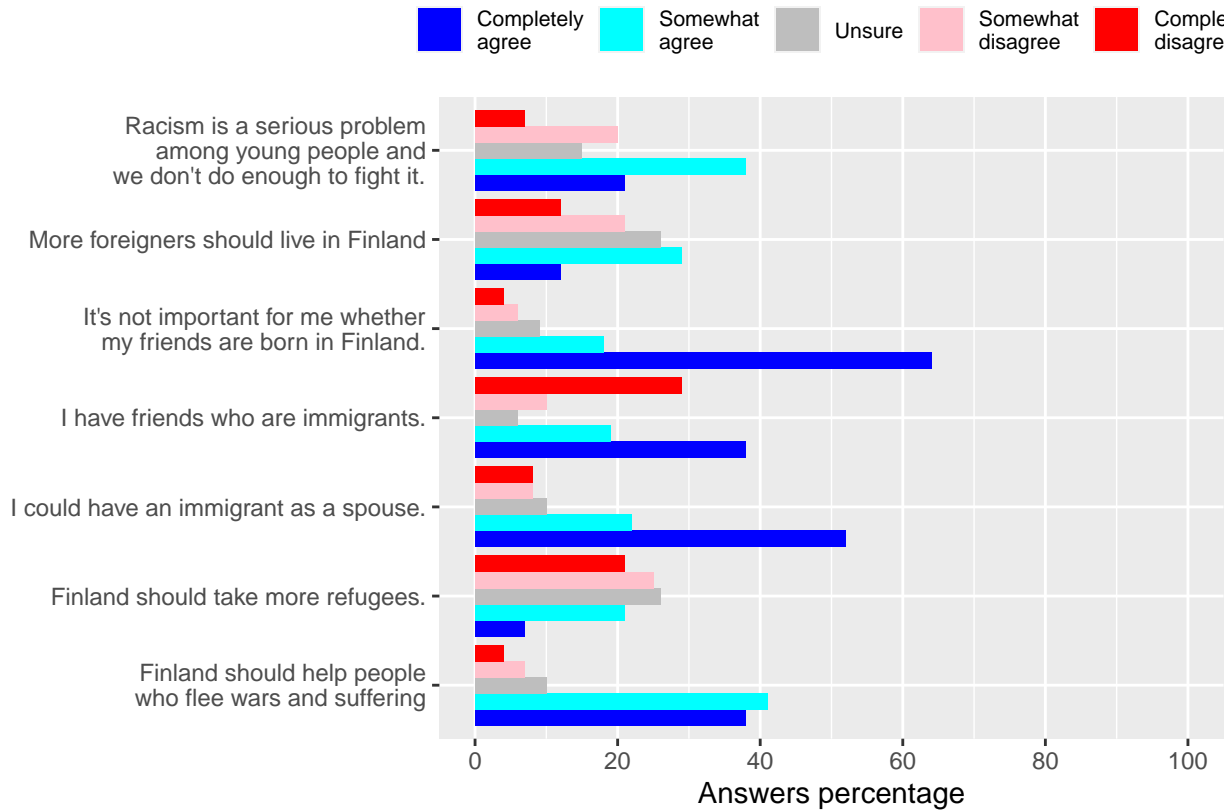
```
library(dplyr)
library(tidyr)
library(ggplot2)

dane <- read.csv("E:/RStudio/eskplo/Jakub_Kubacki_PD2/dane_do_wykresu.csv", sep = ';')
# Wiem, że nieelegancko, ale to jedyny sposób jaki mi przyszedł do głowy na
# wrzucenie enterów, żeby się pytania dobrze wyświetlały.
labelki <- c("Finland should help people\nwho flee wars and suffering",
             "Racism is a serious problem\namong young people and
             we don't do enough to fight it.",
             "I have friends who are immigrants.",
             "More foreigners should live in Finland",
             "Finland should take more refugees.",
             "I could have an immigrant as a spouse.",
             "It's not important for me whether\nmy friends are born in Finland.")

# Na potrzeby ggplota potrzebuję, żeby ramka była w postaci wąskiej a kolumny
# tekstowe były factorami.
dane %>%
  mutate(Question = labelki) %>%
  pivot_longer(2:6, names_to = "Answer") -> przetworzone
przetworzone$Answer <- factor(przetworzone$Answer, c("Completely.agree", "Somewhat.agree",
                                                    "Unsure", "Somewhat.disagree",
                                                    "Completely.disagree"),
                             labels = c("Completely\nagree", "Somewhat\nagree",
                                         "Unsure", "Somewhat\ndisagree",
                                         "Completely\ndisagree"))

przetworzone %>%
  ggplot(aes(x = Question, y = value, fill = Answer)) +
  geom_col(position = "dodge") +
  scale_fill_manual(values = c("blue", "cyan", "grey", "pink", "red")) +
  scale_y_continuous(labels = c("0", "20", "40", "60", "80", "100"),
                     breaks = c(0, 20, 40, 60, 80, 100),
                     limits = c(0, 100)) +
  labs(x = "", y = "Answers percentage", fill = "",
       title = "Views on racism and immigration in Finland as of 2014") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(legend.position = "top", legend.text = element_text(size = 8)) +
  theme(plot.margin = grid::unit(c(0, 0, 0, 0), "mm")) +
  expand_limits(y = 150) +
  coord_flip()
```

Views on racism and immigration in Finland as of 201



Jakie zmiany wprowadziłem i dlaczego?

Po pierwsze - poprawiłem typ wykresu. Zgodnie z tym, co było przedstawione na wykładzie, ludzie łatwiej porównują wysokości słupków niż ich proporcje. Danych było mało, więc nic nie stało na przeszkodzie by użyć wykresu słupkowego. Zmiana typu wykresu pozwoliła mi też pozbyć się liczbowych podpisów na wykresie, co pozbywa się problemu procentów nie sumujących się do 100 (oraz minimalizuje liczbę elementów na wykresie).

Po drugie - poprawiłem schemat kolorów. Oryginalny, nie dość, że jest nieczytelny przez zlewanie się kolorów, to jeszcze jest niekonsystentny, bo czasem ksenofobiczna odpowiedź jest kolorowana najciemniej, a czasem najjaśniej.

Dlatego zaprzeczyłem dwóm pytaniom i zastosowałem paletę dwu-gradientową, która wyraźnie podkreśla, po której stronie spektrum politycznego dana grupa jest.