

HW3

Igor Lechoszest

3 kwietnia 2024

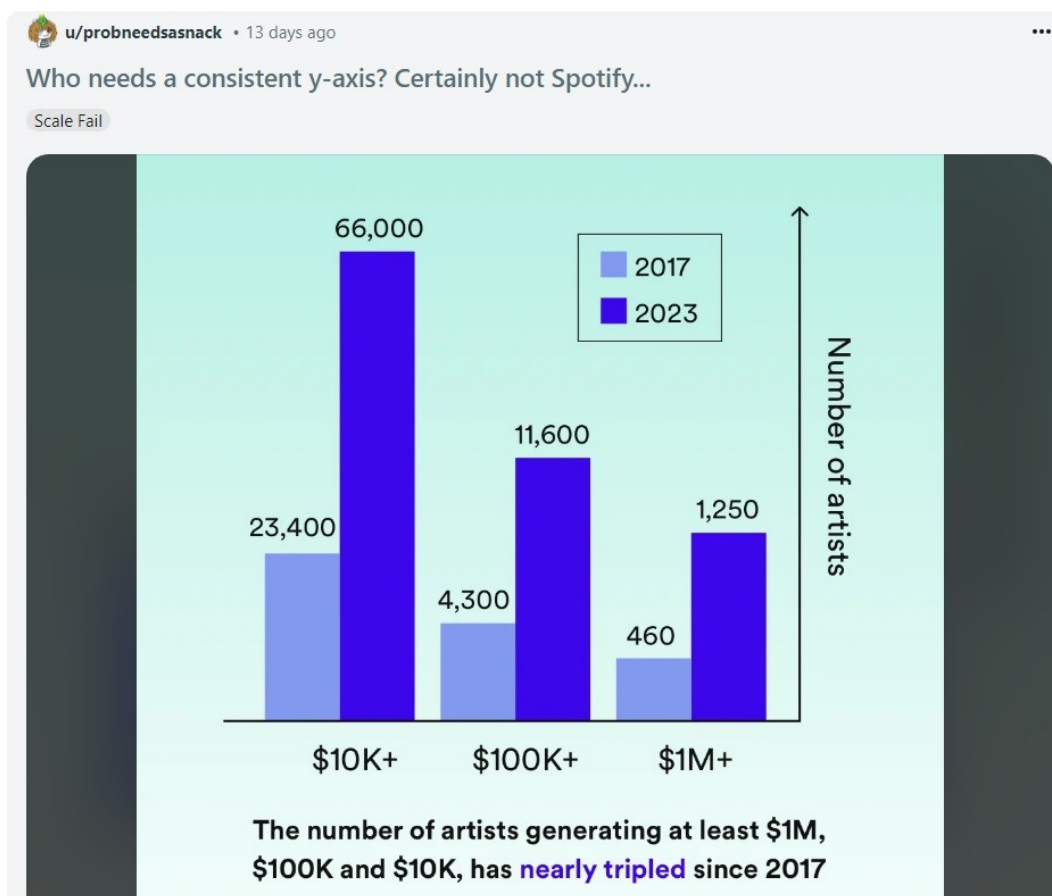
1 Wizualizacja źródłowa

Data publikacji: 20.03.2024

Źródło: Reddit

Link: https://www.reddit.com/r/dataisugly/comments/1bjlgdc/who_needs_a_consistent_yaxis_certainly_not_spotify/

Na znalezionym przeze mnie wykresie widzimy, że oś y praktycznie nie ma skali. 1250 jest wyżej od 23400 ale niżej od 11600. Widać, że nie skala jest najważniejsza, ponieważ wykres ma obrazować to, że w każdej kategorii potroiła się liczba wykonawców. Można to jednak zrobić lepiej nie zakłamując przy tym rzeczywistości.

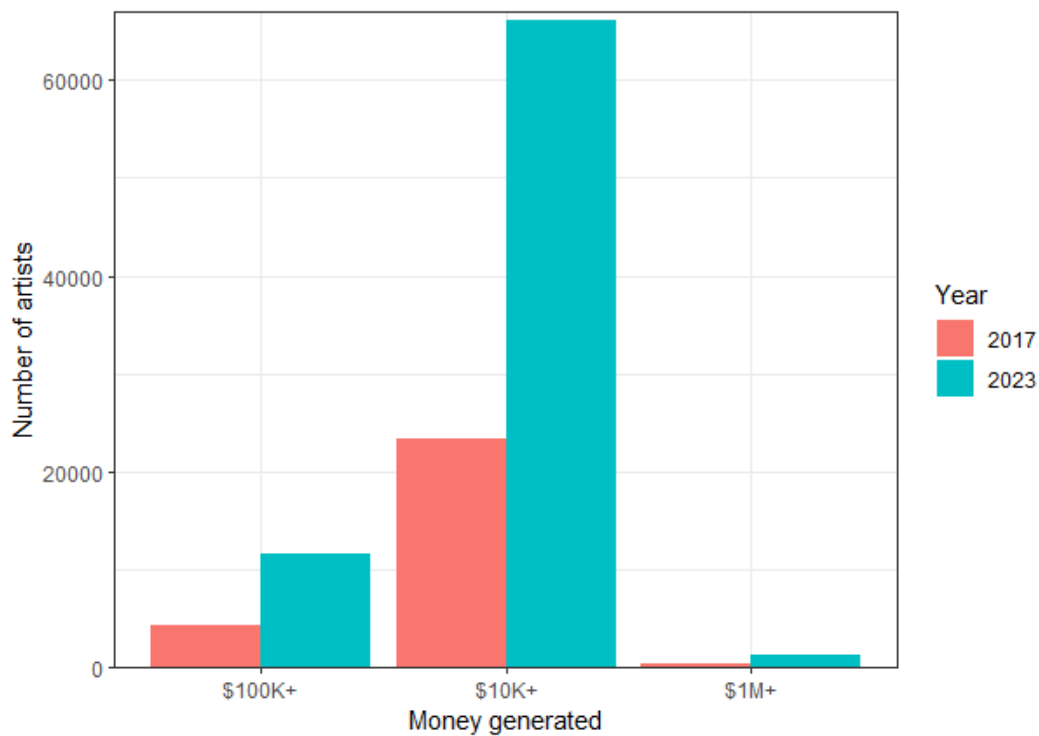


Rysunek 1: Tak się nie robi wykresów

2 Poprawiona wizualizacja

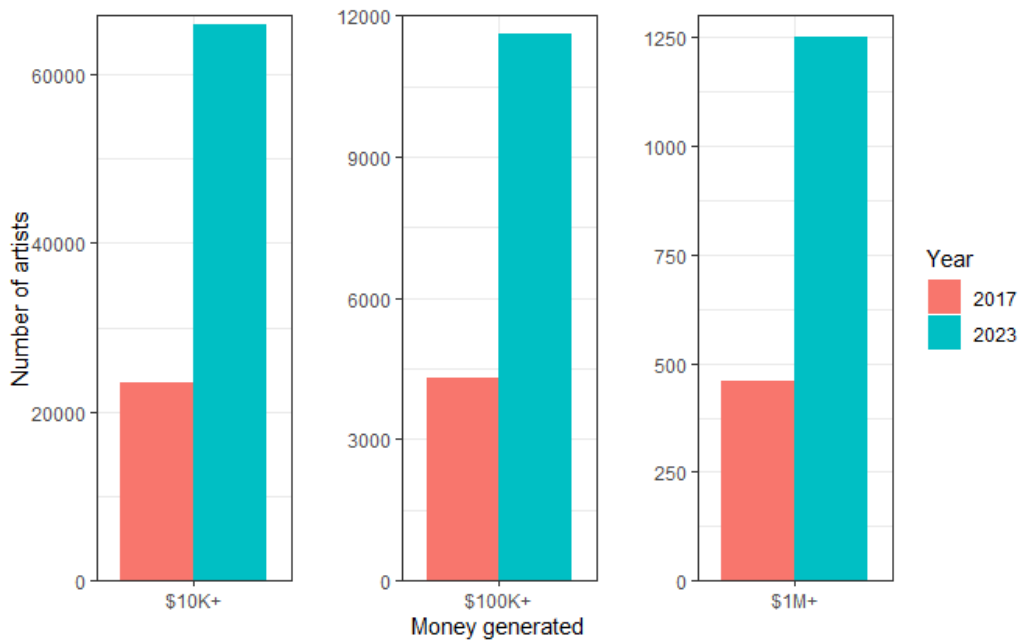
Postanowiłem poprawić tę wizualizację na dwa sposoby. Jeden, moim zdaniem gorszy, obrazuje dane w sposób bezwzględny. Drugi jest bardziej zbliżony do materiału źródłowego, ale nie widać na nim tak dobrze bezwzględnych różnic pomiędzy liczbą artystów w danych kategoriach.

Wykres 1:



Rysunek 2: Może i przedstawia prawdziwe dane, ale mało na nim widać

Wykres 2:



Rysunek 3: Tu dobrze widać, że w każdej kategorii potroiliła się liczba artystów

Oba wykresy przedstawiają prawdziwe informacje w przeciwieństwie do wykresu źródłowego. Wystraczyło zadbać o poprawną skalę na osi y.

3 Kod generujący wykresy

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

df1 <- data.frame(kasa = "$10K+", rok2017 = 23400, rok2023 = 66000)
df2 <- data.frame(kasa = "$100K+", rok2017 = 4300, rok2023 = 11600)
df3 <- data.frame(kasa = "$1M+", rok2017 = 460, rok2023 = 1250)
colnames(df1) <- c("kasa", "2017", "2023")
colnames(df2) <- c("kasa", "2017", "2023")
colnames(df3) <- c("kasa", "2017", "2023")
w1 <- df1 %>%
  pivot_longer("2017":"2023") %>%
  ggplot()+
  geom_col(aes(x = kasa, y = value, fill = name), position = "dodge") +
  labs(x = "", y = "Number of artists")+
  theme_bw()+
  theme(legend.position = "none")+
  scale_y_continuous(expand = c(0,0), limits = c(0,67000))
w2 <- df2 %>%
  pivot_longer("2017":"2023") %>%
  ggplot()+
  geom_col(aes(x = kasa, y = value, fill = name), position = "dodge") +
  labs(x = "Money generated", y = "") +
  theme_bw()+
```

```

    theme(legend.position = "none")+
    scale_y_continuous(expand = c(0,0), limits = c(0,12000))
w3 <- df3 %>%
  pivot_longer("2017":"2023") %>%
  ggplot()+
  geom_col(aes(x = kasa, y = value, fill = name), position = "dodge")+
  labs(x = "", y = "", fill = "Year")+
  theme_bw()+
  scale_y_continuous(expand = c(0,0), limits = c(0,1300))

w1+w2+w3+ plot_layout(guides = "collect")

df <- data.frame(kasa = c("$10K+", "$100K+", "$1M+"), rok2017 = c(23400, 4300, 460), rok2023 = c(66000,

colnames(df) <- c("kasa", "2017", "2023")
w4 <- df %>%
  pivot_longer("2017":"2023") %>%
  ggplot() +
  geom_col(aes(x = kasa, y = value, fill = name), position = "dodge") +
  labs(x = "Money generated", y = "Number of artists", fill = "Year") +
  theme_bw()+
  scale_y_continuous(expand = c(0,0), limits = c(0,67000))
w4

```