

# Wstęp do eksploracji danych, Praca domowa 2

Jan Pogłód

March 2023

## 1 Wstęp

W tej pracy domowej postawiłem sobie za zadanie zbadanie umiejętności piłkarzy z poszczególnych krajów na mistrzostwach świata w piłce nożnej w latach 1930-2022. FIFA w każdym z następujących po sobie mistrzostw świata wybiera najlepszych piłkarzy zwarzając na ich umiejętności i w ten sposób przydziela co roku kolejne nagrody: Golden Boot, Golden Glove, Silver Boot, Silver Ball, Bronze Boot, Bronze Ball, Best Young Player. W moich kolejnych krokach i wizualizacji badać będę które z krajów dostały takich nagród najwięcej



### 1.1 Źródło danych:

[https://github.com/jfjelstul/worldcup/blob/master/data-csv/award\\_w\\_inners.csv](https://github.com/jfjelstul/worldcup/blob/master/data-csv/award_w_inners.csv)

## 2 Selekcja i obróbka danych

### 2.1 Generowanie danych

Zaczniemy tworzenie naszego wykresu od zaimportowania potrzebnych bibliotek i danych

```
library(dplyr)
library(ggplot2)
library(maps)
dane <- read.csv("award_winners.csv")
world <- map_data("world")
```

Następnym krokiem będzie znalezienie Krajów na świecie, które dostały przynajmniej jedną nagrodę na okresie wszystkich turniejów. Te kraje to:

```
Kraje <- as.list(unique(dane %>% select(team_name)))
Kraje_z_danych <- c(
  "Argentina", "Uruguay", "United_States", "Czechia",
  "Germany", "Italy", "Brazil", "Hungary",
  "Spain", "Switzerland", "West_Germany", "Austria",
  "France", "Soviet_Union", "Yugoslavia", "Chile",
  "Portugal", "England", "Peru", "Poland",
  "Netherlands", "Denmark", "Belgium", "Cameroon",
  "Bulgaria", "Russia", "Sweden", "Croatia",
  "South_Korea", "Colombia")
```

## 2.2 Obróbka danych:

Zliczanie interesujących mnie wyników wykonuje po krajach które dostawały nagrody na turnieju. W ten sposób otrzymuję dokładną ilość otrzymanych nagród dla piłkarzy z poszczególnych krajów.

```
dane_swiat <- dane %>% select(team_name,award_name) %>%  
  filter(team_name %in% Kraje_z_danych) %>%  
  group_by(team_name) %>%  
  summarize(count=n()) %>%  
  mutate(region=team_name)  
  
merge_swiat <- full_join(world, dane_swiat, by="region") %>%  
  mutate(count = ifelse(is.na(count), 0, count))
```

## 2.3 Tworzenie mapy za pomocą pakietu ggplot2 i map:

```
mapa1 <- ggplot(merge_swiat, mapping=aes(x=long,y=lat,group=group,  
fill=count)) +  
  geom_polygon(color="white") +  
  scale_fill_gradient2(low="#F0ECCF",  
                        mid="#61876E",  
                        high="#3C6255",  
                        midpoint=10) + theme_minimal() +  
  coord_fixed(1.3) +  
  labs(title = "Piłkarze z różnych krajów dostawali  
najwięcej nagród  
za osiągnięcia mistrzostwach świata?",  
        subtitle = "W latach 1930–2022",  
        x="długość geograficzna",y="szerokość geograficzna",  
        fill="Ilość nagród:") +  
  theme(legend.position = "bottom")
```

## 2.4 Analogicznie postępujemy tworząc identyczną mapę, ale tylko dla Europy:

```
Kraje_europa <- c(
  "Germany", "Italy", "Hungary",
  "Spain", "Switzerland", "Austria",
  "France", "Portugal", "England", "Poland",
  "Netherlands", "Denmark", "Belgium",
  "Bulgaria", "Sweden", "Croatia")

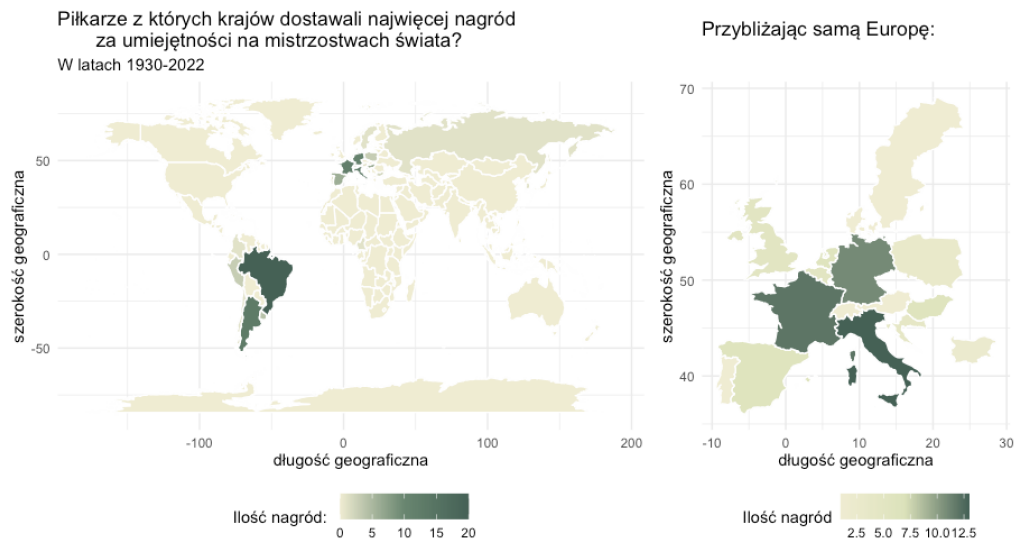
dane_europa <- dane %>%
  select(team_name, award_name) %>%
  filter(team_name %in% Kraje3) %>%
  group_by(team_name) %>%
  summarize(count=n()) %>%
  mutate(region=case_when(team_name=="England" ~ "UK",
                           team_name!="England" ~ team_name))

merge_europa <- inner_join(world, dane_europa, by="region") %>%
  mutate(count = ifelse(is.na(count), 0, count))
```

## 2.5 Tworzymy mapę dla Europy, aby nasz czytelnik miał czytelniejsze wyniki

```
wykres2 <- ggplot(merge_europa, mapping=aes(x=long, y=lat, group=group,
fill=count)) +
  geom_polygon(color="white") +
  scale_fill_gradient2(low="#F0ECCF", mid="#DAE2B6",
                      high="#3C6255", midpoint=7) +
  theme_minimal() +
  coord_fixed(1.3) +
  labs(title = "Przybli aj c sam u Europ :",
       x="d ugo u geograficzna", y="szeroko u geograficzna",
       fill="Ilo u nagr d") +
  theme(legend.position = "bottom")
library(patchwork)
wykres1+wykres2
```

## 2.6 Wizualizacja wyniku:



## 3 Podsumowanie

Za pomocą naszej czytelnej mapy widzimy, że to Brazylia i Włochy mają najzdolniejszych piłkarzy. To te kraje spośród pozostałych najbardziej rozróżniają się pod względem zdobytych nagród za umiejętności piłkarzy. Oznacza to, że bardzo prawdopodobne, że urodzenie się w tych właśnie krajach może wpłynąć na rozwój kariery i talentów piłkarskich.