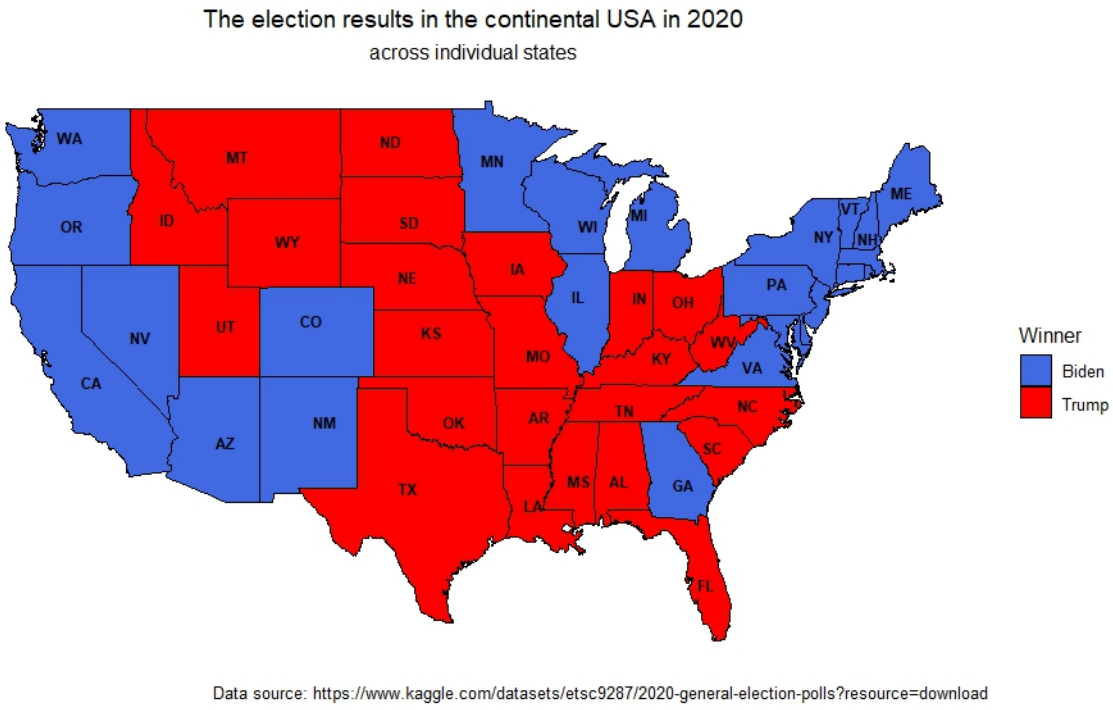
# Praca domowa 4 z eksploracji danych

Maciej Turczyński

## 1 Utworzona wizualizacja




Rysunek 1: Utworzona wizualizacja

Źródło danych o wyborach: `https://www.kaggle.com/datasets/etsc9287/2020-general-election-polls?resource=download`

## 2 Opis wykresu

Utworzony przeze mnie wykres przedstawia wyniki wyborów w kontynentalnych Stanach Zjednoczonych Ameryki. Te pomalowane na czerwono przedstawiają stany, w których triumfował Donald Trump, zaś na niebiesko - Joe Biden.

## 3 Kod do wygenerowania wizualizacji

```
library(ggplot2)
library(maps)
library(mapdata)
library(dplyr)
library(leaflet)
raw_data <- read.csv('county_statistics.csv')
data <- raw_data %>%
  mutate(
    Donald_Trump_votes = as.numeric(votes20_Donald_Trump),
    Joe_Biden_votes = as.numeric(votes20_Joe_Biden),
    total_votes = as.numeric(total_votes20)
  ) %>%
  select(county,
         state,
         Donald_Trump_votes,
         Joe_Biden_votes,
         total_votes) %>%
  group_by(state) %>%
  summarise(
    Trump_percent = sum(Donald_Trump_votes, na.rm = TRUE) / sum(
       total_votes, na.rm =
                                                                TRUE
                                                                   )
                                                                   *
                                                                   100,
    Biden_percent = sum(Joe_Biden_votes, na.rm = TRUE) / sum(total_
       votes, na.rm =
                                                            TRUE)
                                                                *
                                                                100
  )
```

```
colnames(data) <- c('Shortcut', 'Trump_percent', 'Biden_percent')
states <- read.csv("states.csv")
colnames(states) <- c('State', 'Shortcut')

final_data <- left_join(data, states, by = 'Shortcut') %>%
  select(State, Biden_percent, Trump_percent)

final_data$State <- tolower(final_data$State)

usa <- map_data("state")

map_data <- left_join(usa, final_data, by = c('region' = 'State'))

map_data <- map_data %>%
  mutate(winner = ifelse(Biden_percent > Trump_percent, "Biden", "
     Trump"))

states_mid <- read.csv('states_names.csv')
state_middle <- states_mid %>%
  slice(-c(1,7,8,9,12,20,21,32,40,41))

ggplot() +
  geom_map(data = map_data, map = map_data,
           aes(x = long, y = lat, map_id = region, fill = winner),
           color = "black", size = 0.15) +
  expand_limits(x = map_data$long, y = map_data$lat) +
  labs(title = "The election results in the continental USA in 2020"
     ,
       subtitle = "across individual states",
       fill = "Winner") +
  theme_minimal() +
  theme(legend.position = "right") +
  scale_fill_manual(values = c("Biden" = "royalblue", "Trump" = "red
     ")) +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  geom_text(data = state_middle, aes(x = longitude, y = latitude,
     label = state),
            size = 3, fontface = "bold", color = "black") +
```

```
66    labs(caption = "Data source: https://www.kaggle.com/datasets/
          etsc9287/2020-general-election-polls?resource=download")
```