# Zadanie 1. Analiza pojedynczego zdania ----

# Wczytaj dane tekstowe

text <- "And so even though we face the difficulties of today and tomorrow, I still have a dream."

text

# Sprawdź częstości słów za pomocą pakietu qdap

install.packages("qdap")

library(qdap)

freq\_terms(text)

# Zapisz najczęściej występujące terminy w ramce danych

frequent\_terms <- freq\_terms(text)

frequent\_terms

# Wizualizacja najczęściej występujących terminów

plot(frequent\_terms)

# UWAGA

# Słowa nie są wymienione w takiej kolejności, w jakiej występują w zdaniu

# są prezentowane w porządku alfabetycznym.

# Takie podejście nazywa się Bag of Words (torba słów).

# Inne możliwości pakietu qdap

?freq\_terms

# Wizualizacja za pomocą ggplot2

library(ggplot2)

ggplot(frequent\_terms, aes(x = WORD, y = FREQ)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(x = "Słowo", y = "Częstość") +

theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +

ggtitle("Wykres częstości słów")

ggplot(frequent\_terms, aes(y = WORD, x = FREQ)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(x = "Słowo", y = "Częstość") +

theme(axis.text.x = element\_text(angle = 0, hjust = 1)) +

ggtitle("Wykres częstości słów")

# Bardziej atrakcyjna wizualizacja

ggplot(frequent\_terms, aes(x = FREQ, y = reorder(WORD, FREQ))) +

geom\_bar(stat = "identity", fill = "skyblue", color = "darkblue", alpha = 0.8) +

labs(x = "Częstość", y = "Słowo") +

ggtitle("Wykres częstości słów") +

theme\_minimal() +

theme(axis.text.y = element\_text(size = 10), # Dostosowanie rozmiaru czcionki etykiet na osi Y

plot.title = element\_text(hjust = 0.5, size = 16, face = "bold"), # Wyśrodkowanie i stylizacja tytułu wykresu

panel.grid.major.y = element\_blank(), # Usunięcie głównych linii siatki poziomej

panel.grid.minor.y = element\_blank(), # Usunięcie mniejszych linii siatki poziomej

axis.line = element\_line(color = "black")) # Dostosowanie linii osi

# Stopwords (stop słowa – słowa do usunięcia)

# Najczęściej występujące 25, 100 i 200 słów

Top25Words

Top100Words

Top200Words

# Usunięcie stop słów

frequent\_terms2 <- freq\_terms(text, stopwords = Top25Words)

frequent\_terms3 <- freq\_terms(text, stopwords = Top100Words)

frequent\_terms4 <- freq\_terms(text, stopwords = Top200Words)

plot(frequent\_terms4)

# Zadanie 2. Analiza całego akapitu ----

# Wczytaj dane tekstowe

text <- "And so even though we face the difficulties of today and tomorrow, I still have a dream. It is a dream deeply rooted in the American dream."

text

frequent\_terms <- freq\_terms(text)

frequent\_terms

frequent\_terms <- freq\_terms(text, stopwords = Top200Words)

plot(frequent\_terms)