SPRAWOZDANIE

Zajęcia: Analiza Procesów Uczenia Prowadzący: prof. dr hab. Vasyl Martsenyuk

# Laboratorium 7

Data 26.05.2023r.

**Temat:** "Problemy NLP w uczeniu maszynowym"

# Wariant 5

Jarosław Waliczek

Informatyka II stopień,

stacjonarne 1 semestr

# Polecenie:

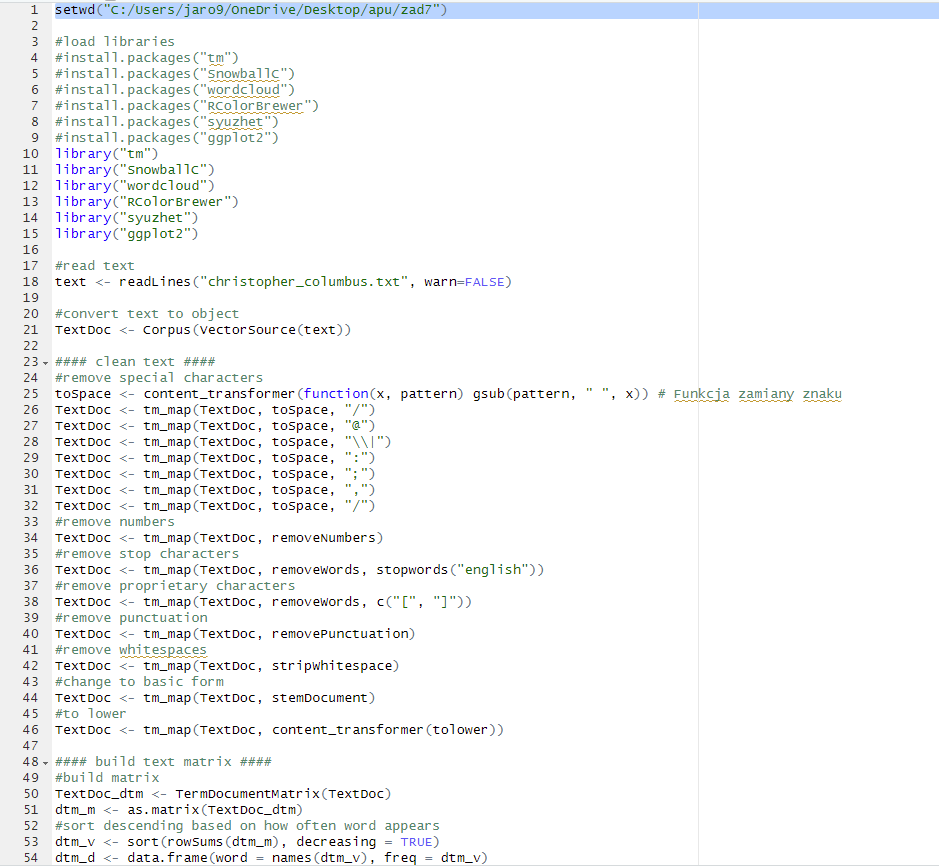
Zadanie dotyczy analizy tekstu, w tym listę częstotliwości słów, budowanie chmury słów, kojarzeń, sentiment analysis, emotion analysis, bigramów, grafów powiązań. Warianty zadania są określone tekstem w języku angielskim umieszczonym na portalu en.wikipedia.org (główna część artykułu bez literatury):

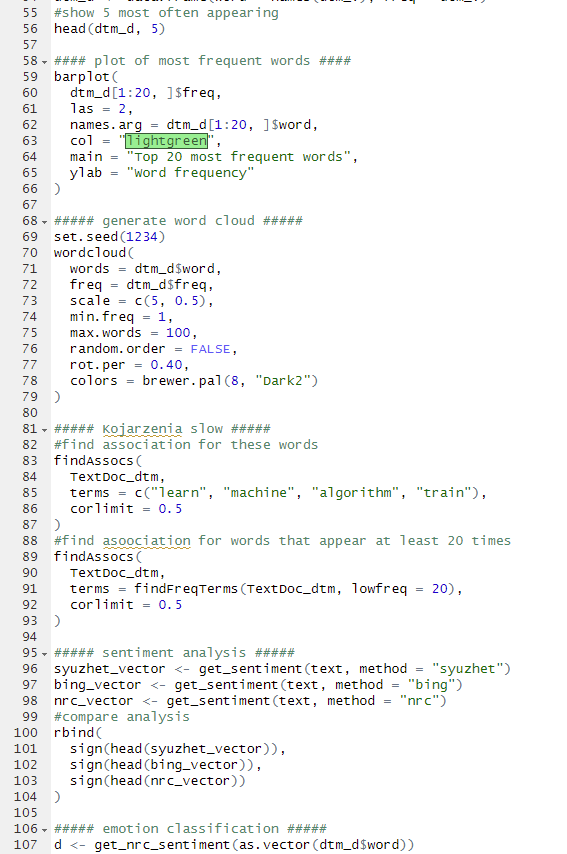
5. <https://en.wikipedia.org/wiki/Christopher_Columbus>

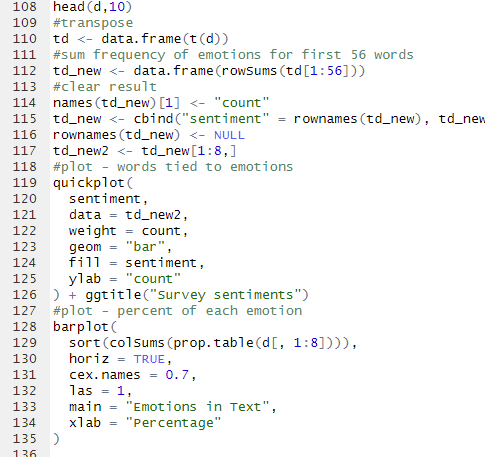
# Wprowadzane dane:

Dane są pobrane ze strony <https://en.wikipedia.org/wiki/Christopher_Columbus>

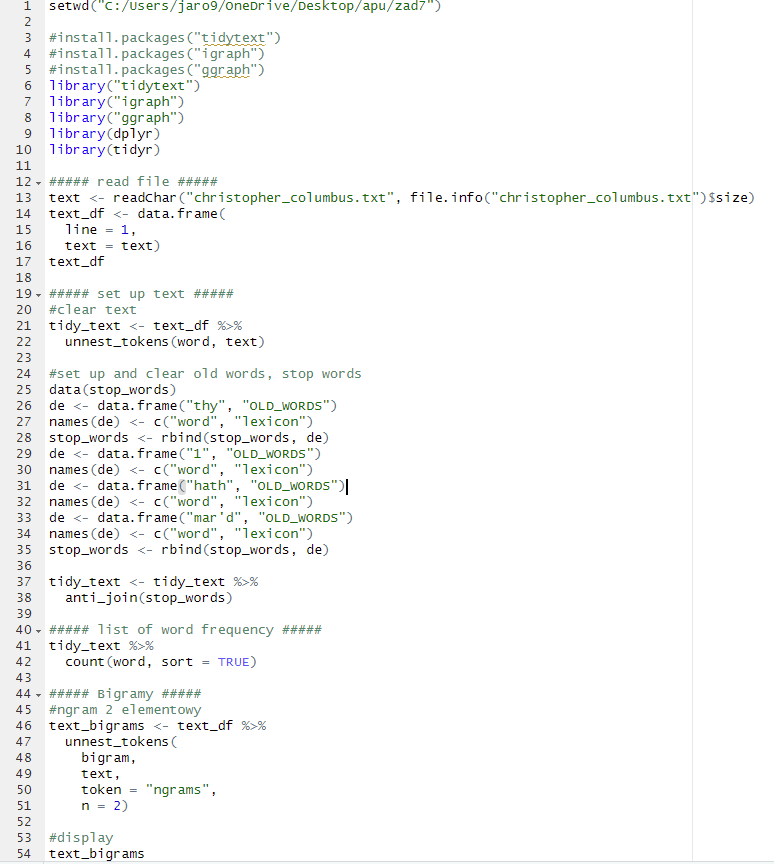
# Wykorzystane komendy:

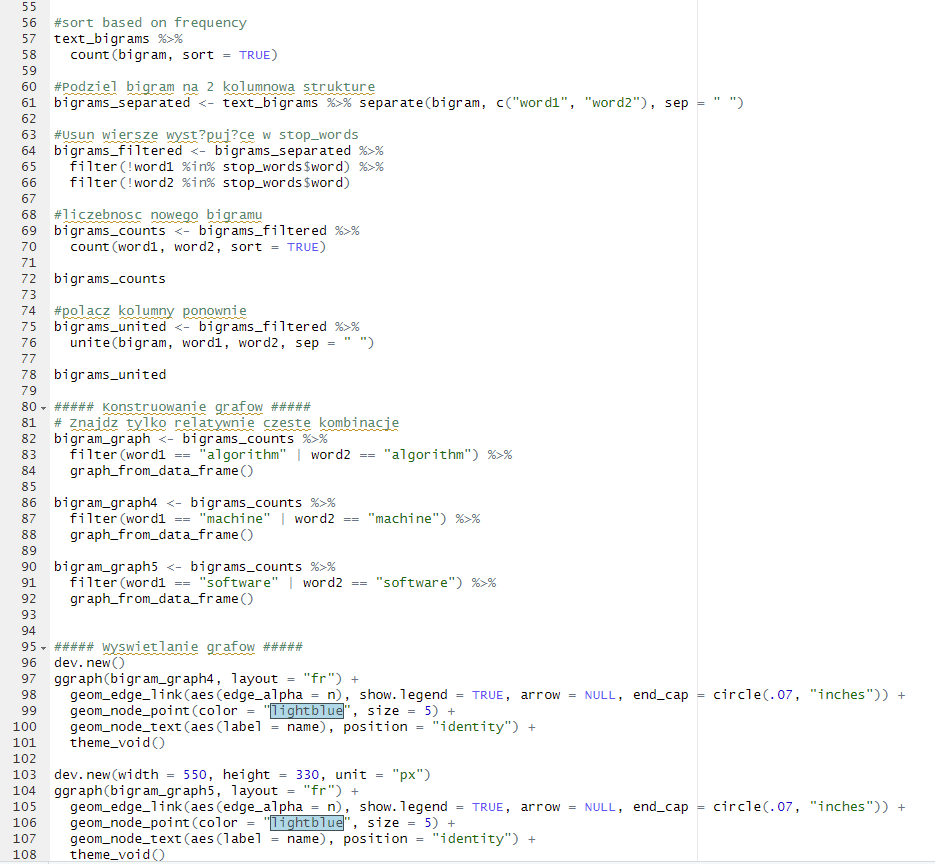
Wstępna analiza tekstów z pomocą list częstotliwości, chmur słów, n-gramów:

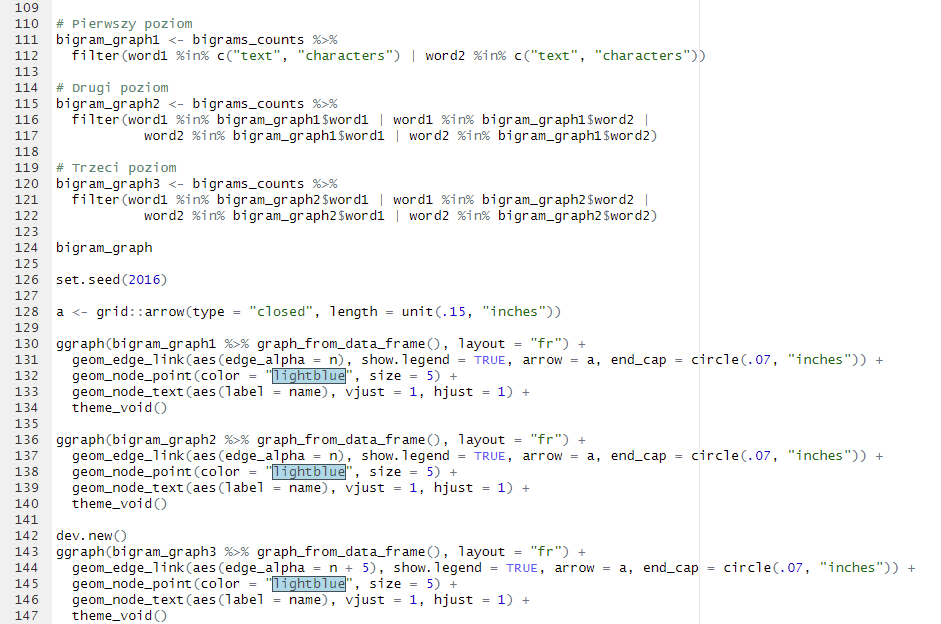




Konstruowanie n-gramów i grafów:

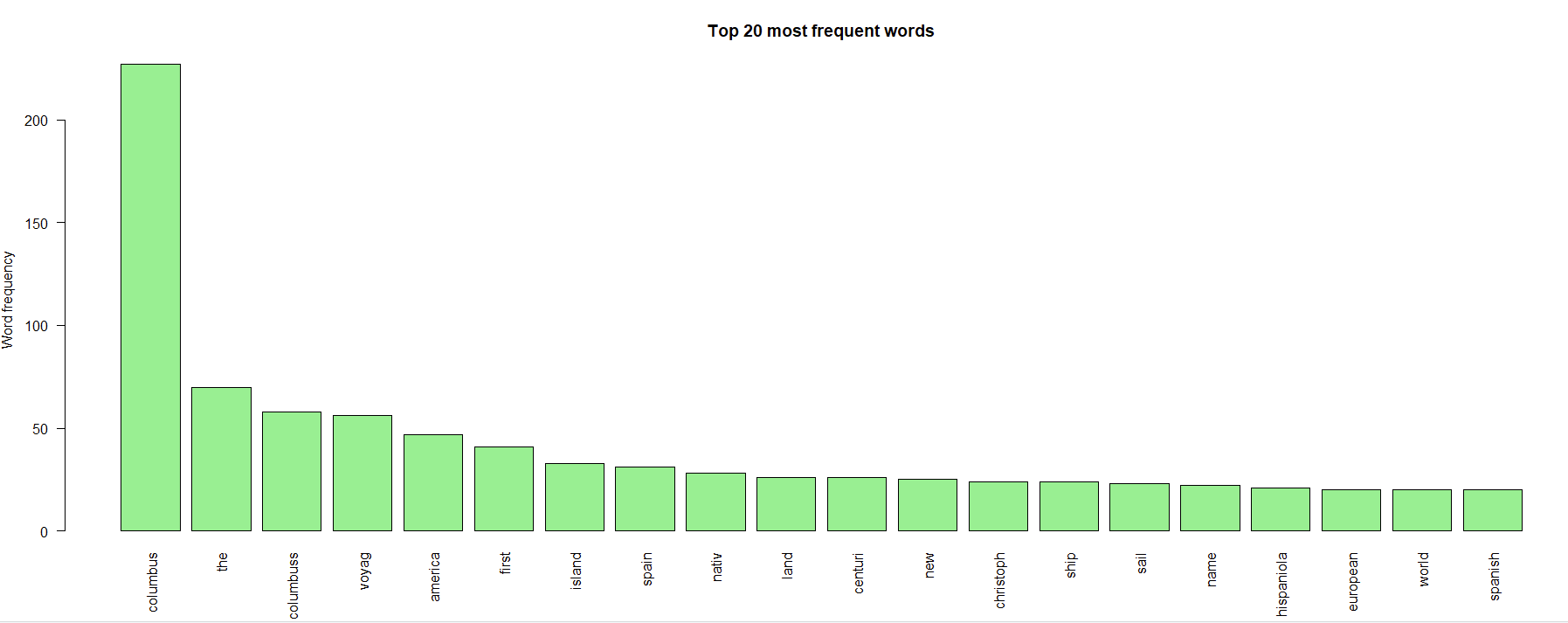


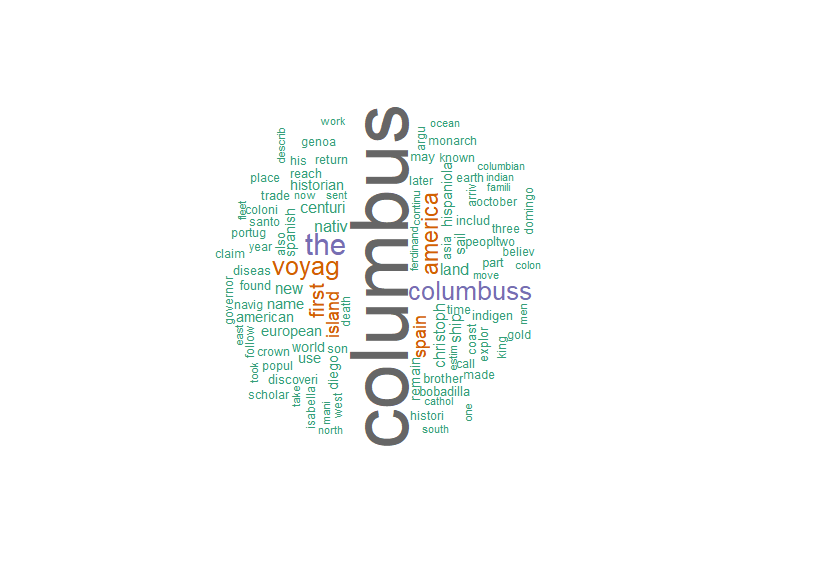


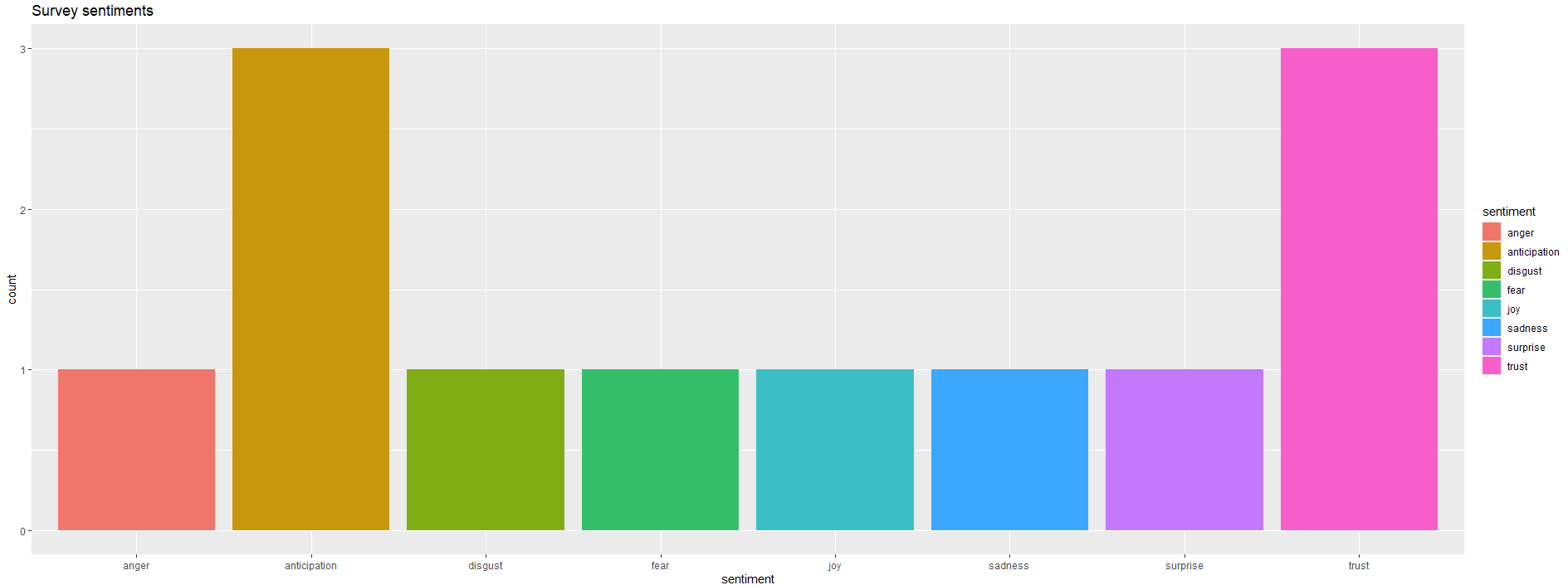
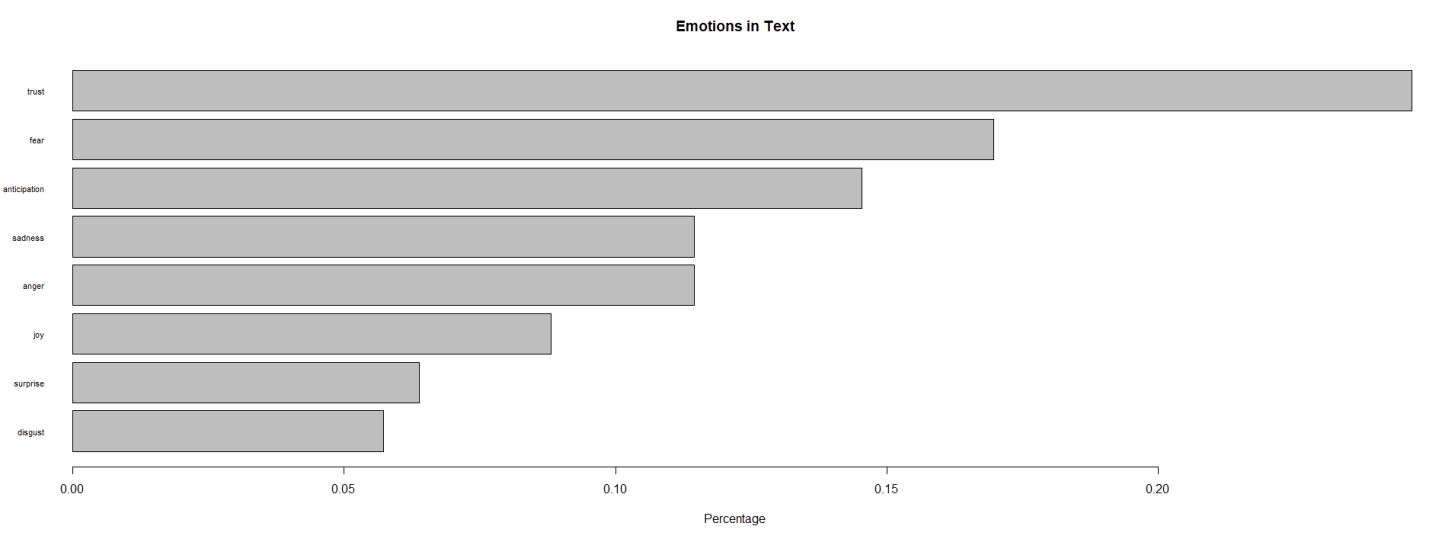


# Wynik działania:

Kod programu dostępny w repozytorium: <https://github.com/Jaro233/APU.git>





# Wnioski:

Dzięki bibliotece Tm, Snowballc, wordcloud itd. możemy dokonać wstępnej analizy tekstu oraz poprzez biblioteki tidytext, igrach, ggraph itp. jesteśmy w stanie konstruować n-gramów i grafów.