



ANALIZA RECENZJI ZE STEAMA

Jakub Bqk

PLAN PREZENTACJI

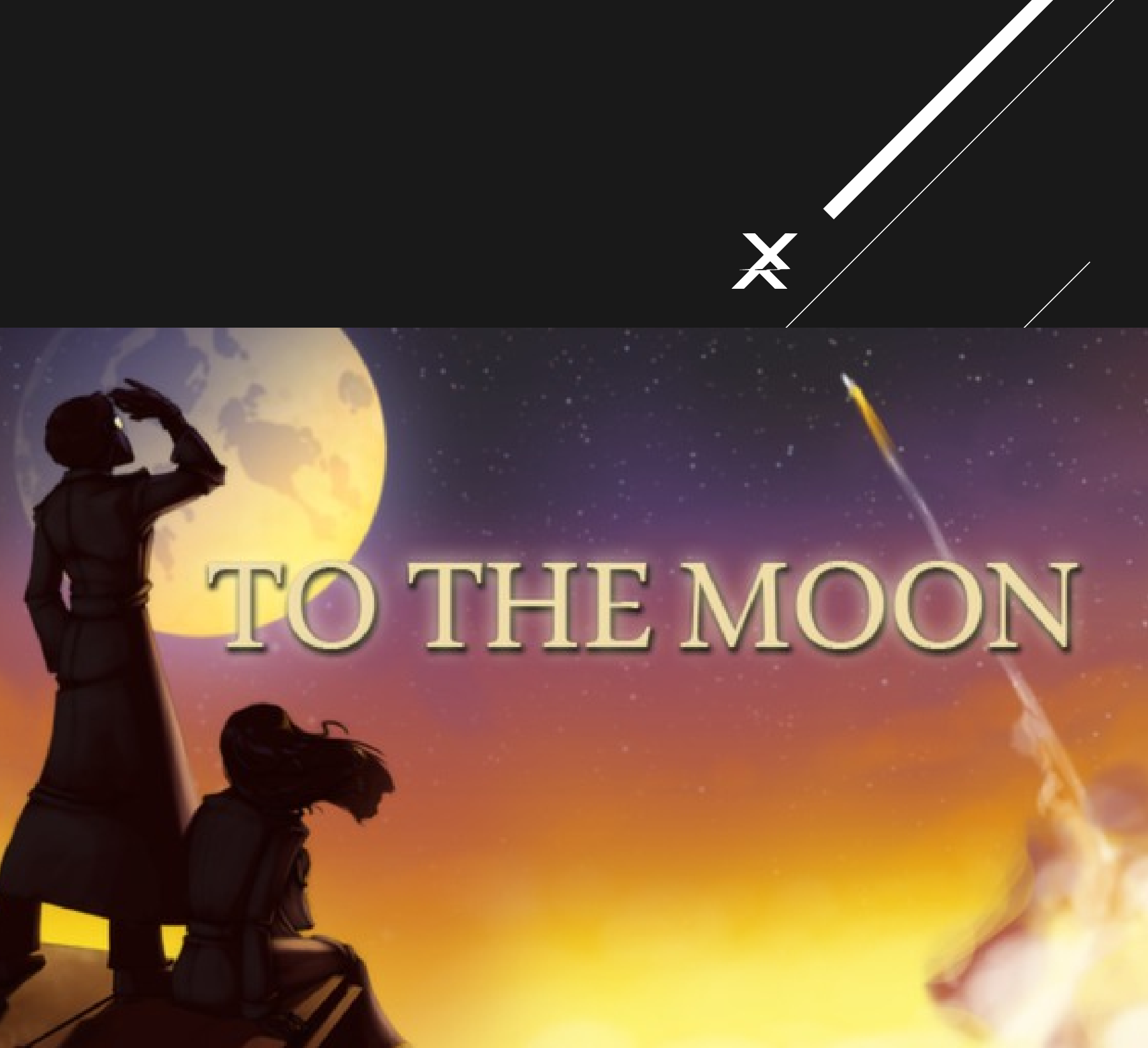
- ▶▶ Zdefiniowany problem
- ▶▶ Parę słów o grze 'To the moon'
- ▶▶ Czynności badawcze
- ▶▶ Podstawowe dane zebrane
- ▶▶ Analiza semantyczna recenzji
- ▶▶ Interesujące fakty
- ▶▶ Podsumowanie



CEL BADAN

**Celem badań było stworzenie
pogłębionej analizy recenzji dla
wybranej gry z popularnej platformy do
kupowania gier Steam.**

W tym celu korzystając z web scrapera pobrałem ponad 2200 recenzji gry 'To the moon', które następnie poddałem analizie.



KILKA INFORMACJI O GRZE

- ▶ ▶ Wydana w 2011 roku przez studio Freebird Games
- ▶ ▶ Gra przygodowa z elementami gry fabularnej
- ▶ ▶ Około 4 godziny potrzebne do przejścia jej
- ▶ ▶ Ponad 96% pozytywnych recenzji na Steamie

CZYNNOŚCI BADAWCZE

- ▶▶ Zebranie danych
- ▶▶ Obróbka i czyszczenie danych
- ▶▶ Podstawowe dane
- ▶▶ Wyodrębnienie najpopularniejszych słów
- ▶▶ Chmura słów
- ▶▶ Sentymenty dla każdego recenzenta
- ▶▶ Wyodrębnienie emocji z recenzji



ANALIZA ZEBRANYCH DANYCH



PODSTAWOWE INFORMACJE

- Zbiór ponad 2300 recenzji zebranych z platformy Steam
- Zebrane dane to m.in. długość czasu spędzonego przez osobę w grze, ilość gier użytkownika, data powstania recenzji czy też sama recenzja
- Zawiera sporo braków(szczególnie w kolumnie 'user_id')



PROCES CZYSZCZENIA DANYCH OBEJMOWAŁ

Usunięcie niepotrzebnych
kolumn oraz braków
danych



Zmiana daty na miesiące
oraz lata



Zmiana części zmiennych
na inny format(najczęściej
'numerical')

PODSUMOWANIE DANYCH

```
> summary(df)
```

recommended	text	hours	username	products
Mode :logical	Length:2330	Min. : 0.000	Length:2330	Min. : 1.0
FALSE:125	Class :character	1st Qu.: 4.100	Class :character	1st Qu.: 123.2
TRUE :2205	Mode :character	Median : 5.000	Mode :character	Median : 277.5
		Mean : 8.805		Mean : 625.9
		3rd Qu.: 6.800		3rd Qu.: 630.2
		Max. :4100.900		Max. :16877.0
		NA's :1		

found_funny	compensation	month	year	word_count
Min. : 0.000	Min. :0.000000	Length:2330	Min. :2012	Min. : 1.00
1st Qu.: 0.000	1st Qu.:0.000000	Class :character	1st Qu.:2015	1st Qu.: 10.00
Median : 0.000	Median :0.000000	Mode :character	Median :2016	Median : 31.00
Mean : 3.017	Mean :0.01974		Mean :2017	Mean : 74.97
3rd Qu.: 1.000	3rd Qu.:0.000000		3rd Qu.:2019	3rd Qu.: 87.00
Max. :425.000	Max. :1.000000		Max. :2021	Max. :1456.00

PODSUMOWANIE DANYCH



**Średnia ilość
godzin na
jednego gracza**

6,5
godziny

**Mediana gier w
bibliotece**

Ponad 277



**Ile % osób
rekomenduje
grę**

94,63 %

**Średnia ilość
zabawnych
reakcji**

9,5

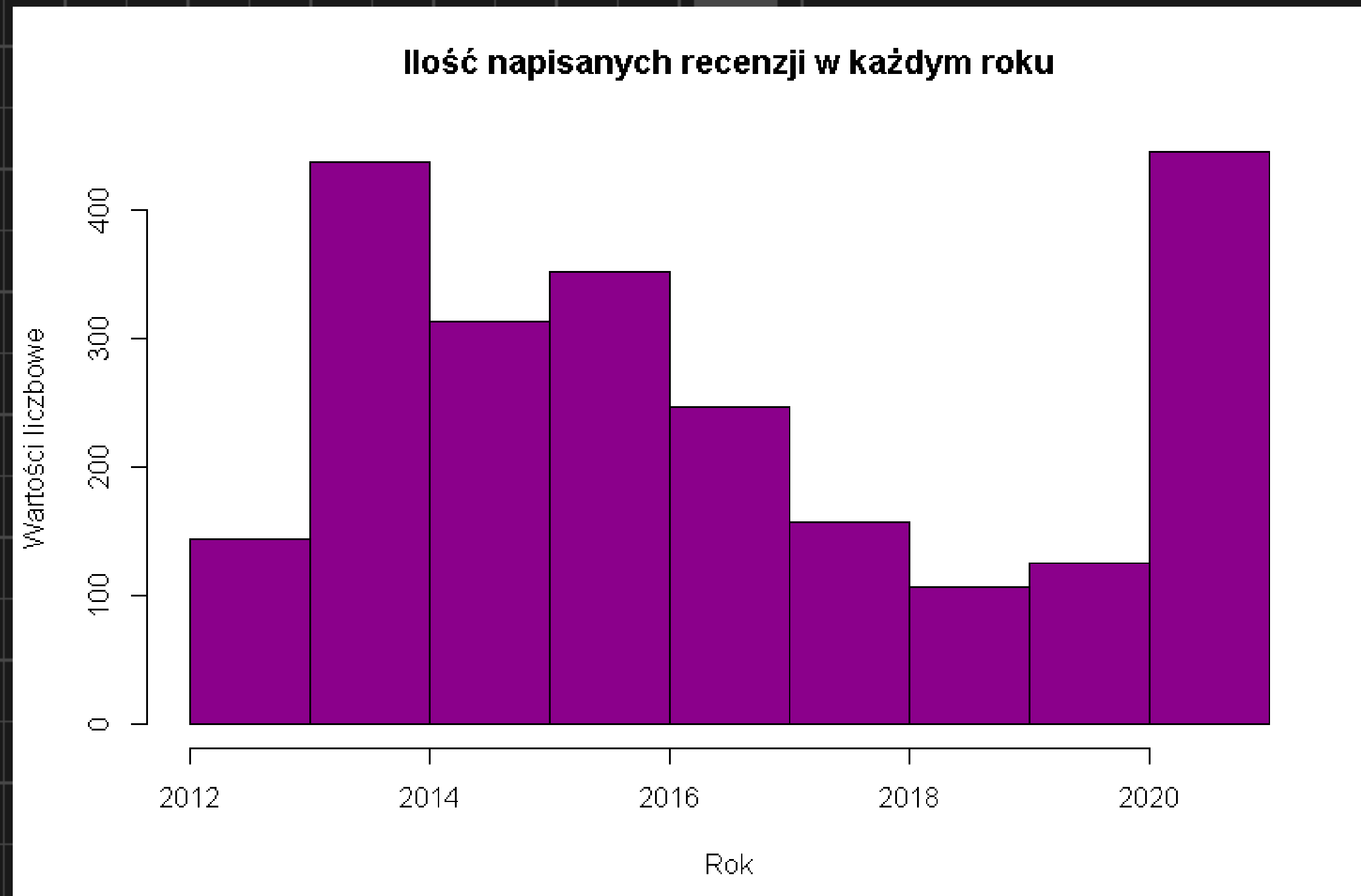


**Średnia
długość
recenzji**

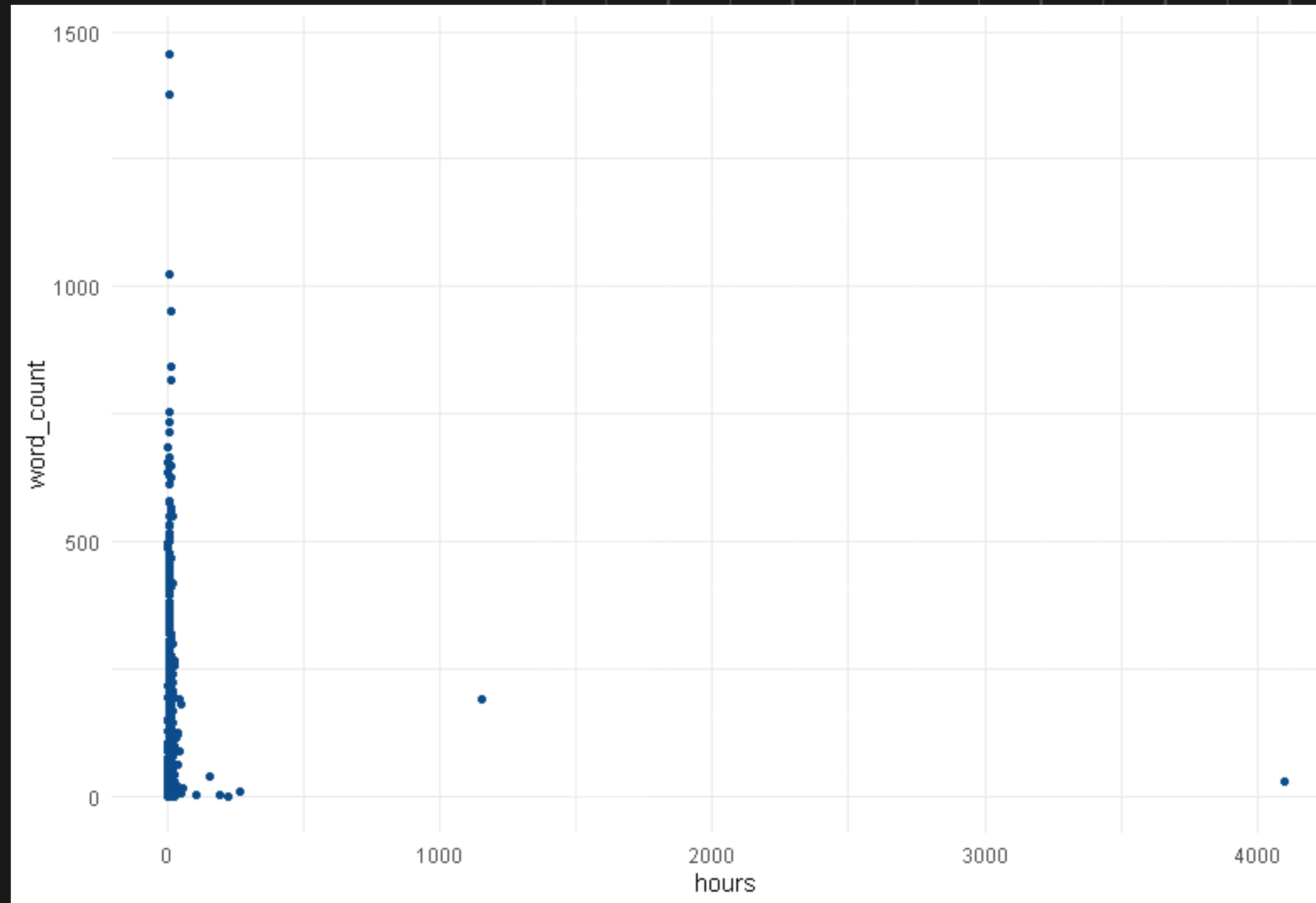
75 słów



ILOSC RECENZJI W ZALEZNOSCI OD ROKU



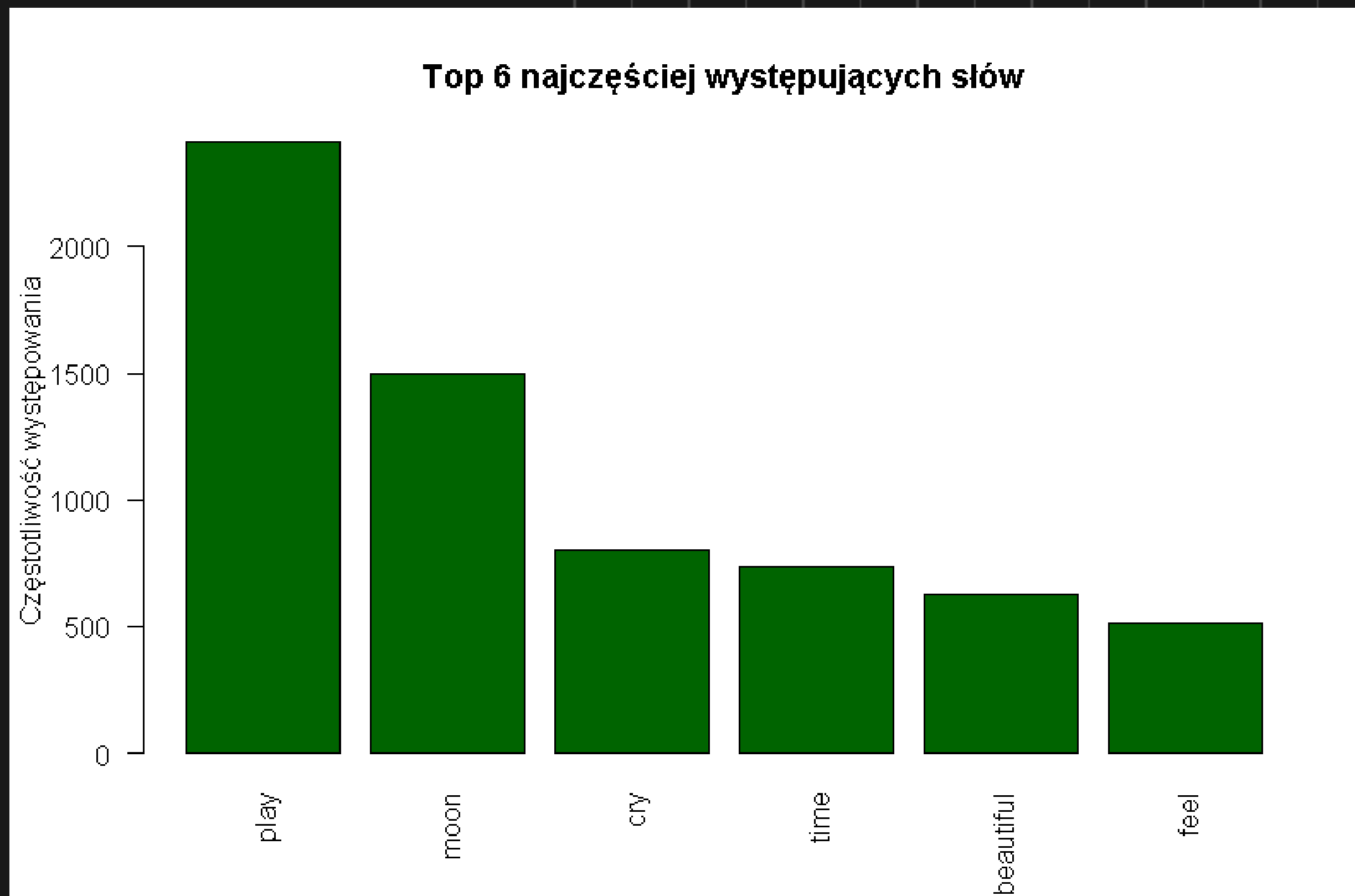
KORELACJA MIĘDZY DŁUGOŚCIĄ RECENZJI, A GODZINAMI SPEDZONYMI W GRZE





ANALIZA SEMANTYCZNA RECENZJI

6 NAJPOPULARNIEJSZYCH SŁÓW





CHMURA SŁÓW

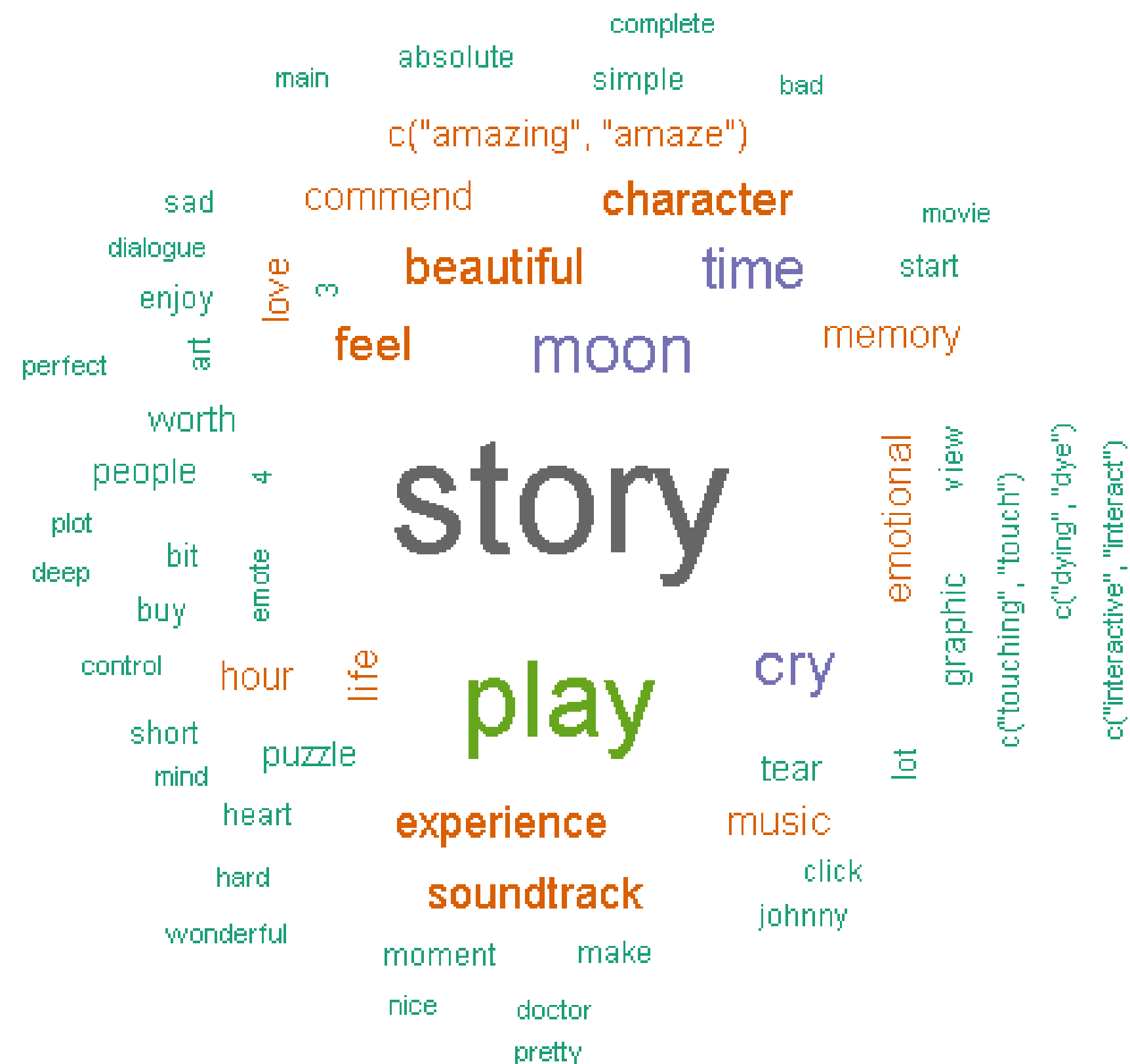


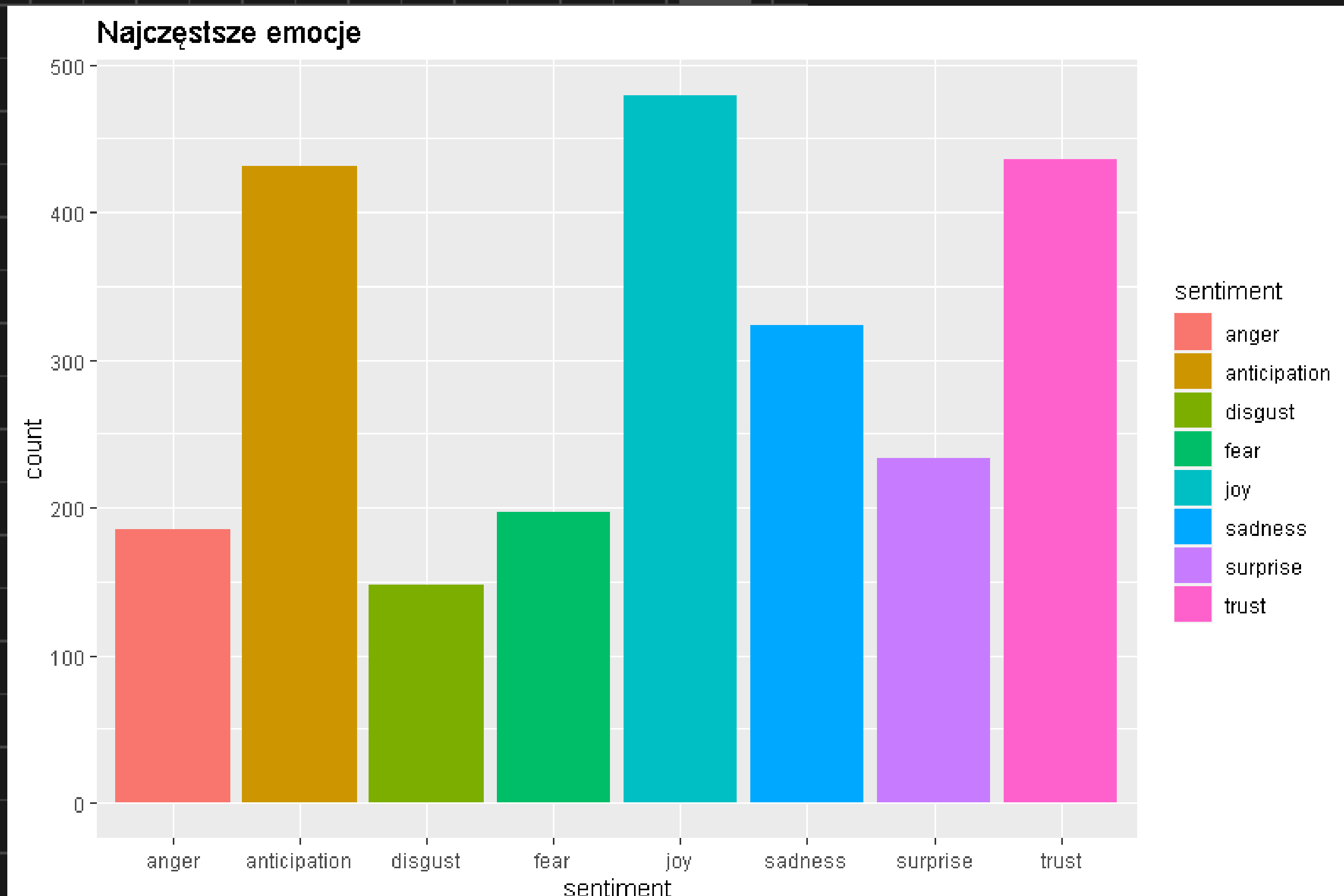


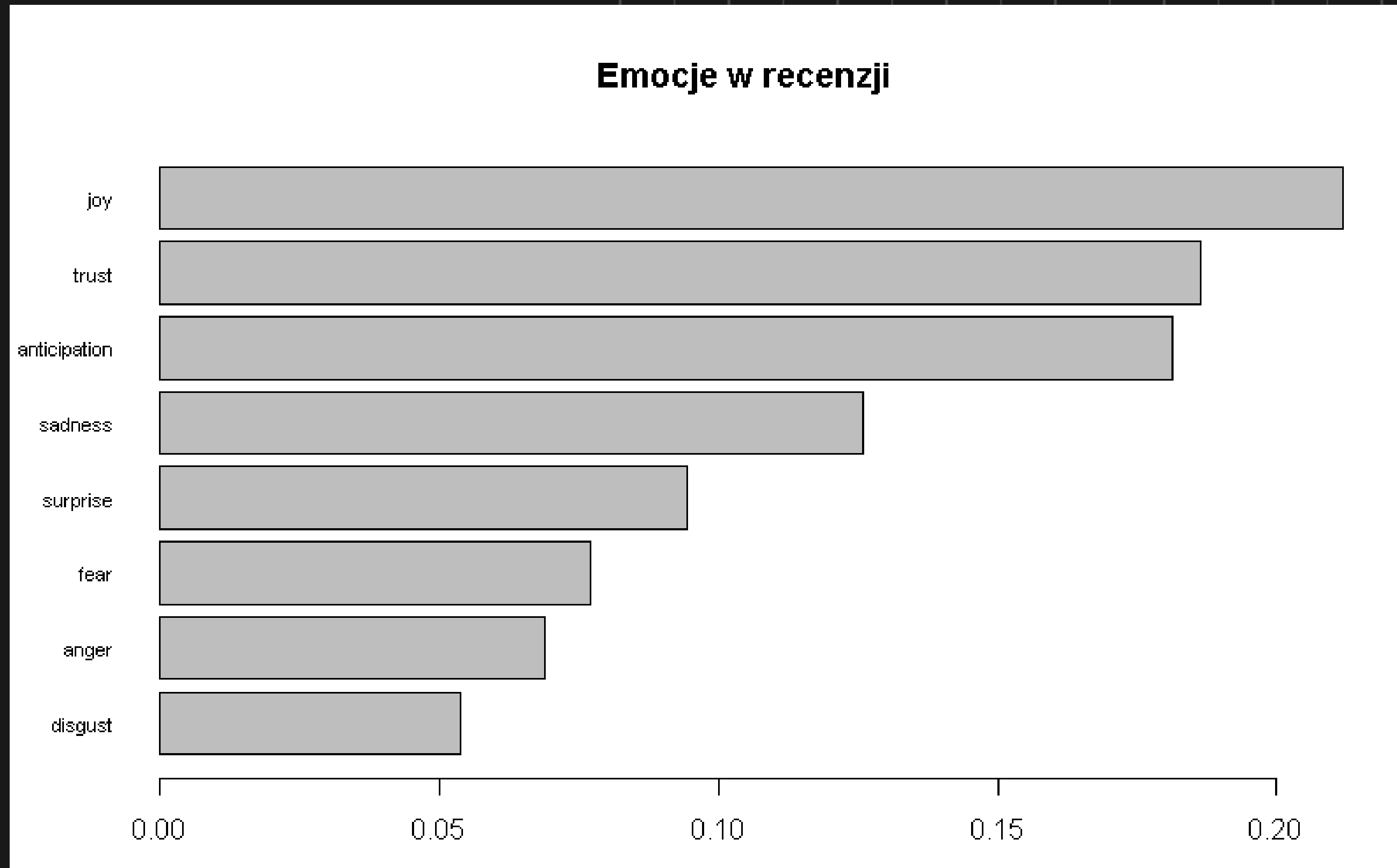
TABELA SENTYMENTÓW

				
Średni sentymen na recenzje	Najbardziej pozytywna recenzja	Najmniej pozytywna recenzja	Modalna wartoŚci sentymen tu	Najczęściej występujący przedział
+ 0,78	+ 22	- 54	1	[-6, 8]
				

* NAJCZĘSCIEJ WYSTĘPUJĄCE EMOCJE



EMOCJE W PROCENTACH



CIEKAWOSTKI

▶▶ **Najdłuższy czas spędzony w grze**

Ponad 4000 godzin, czyli ponad 166 dni

▶▶ **Najdłuższa recenzja**

Ponad 1400 słów, co jest równoważne około 11 stronom A4

▶▶ **Najśmieszniejsza recenzja**

Ma tylko 7 słów, a w całości brzmi: "Dont' cry. Don't cry. Don't cry... I cried"



PODSUMOWANIE

Recenzje gry dobrze oddają nastrój gry

Pozytywne oceny = pozytywny sentyment

Analiza semantyczna może pomóc
twórcom zobaczyć nastawienie graczy do
ich gier

Steam jako dobre miejsce dla Text
Minerów

X

X

BIBLIOGRAFIA

prncc/steam-scraper: A pair of spiders for scraping product data and reviews from Steam. ([github.com](https://github.com/prncc/steam-scraper))

Scraper użyty do pobrania recenzji

Text Mining and Sentiment Analysis: Analysis with R – Simple Talk ([red-gate.com](https://www.simple-talk.com/))



**DZIEKUJE ZA
UWAGE!**

Jakub Bąk