

# Research seminar

## Homework 2

---

BARANOV TIMOFEY

BASB-211



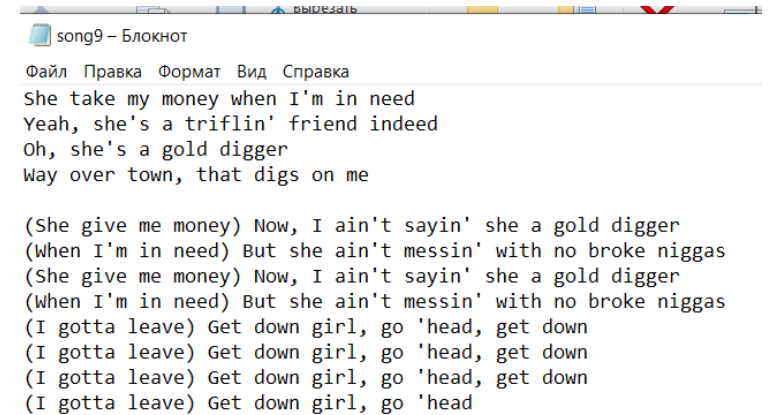
# Data description

Dataset has been collected from the Kanye West's song lyrics.

Top-10 songs has been taken from [www.timeout.com](http://www.timeout.com)

Lyrics from – [www.lyrics.com](http://www.lyrics.com)

Data has been collected manually and stored locally on PC



# Data processing

---

To work with stored data we need to prepare it to analysis, which means:

1. Clean from all punctuations
2. Removing hidden characters
3. Column-based structure

# Popular words

---

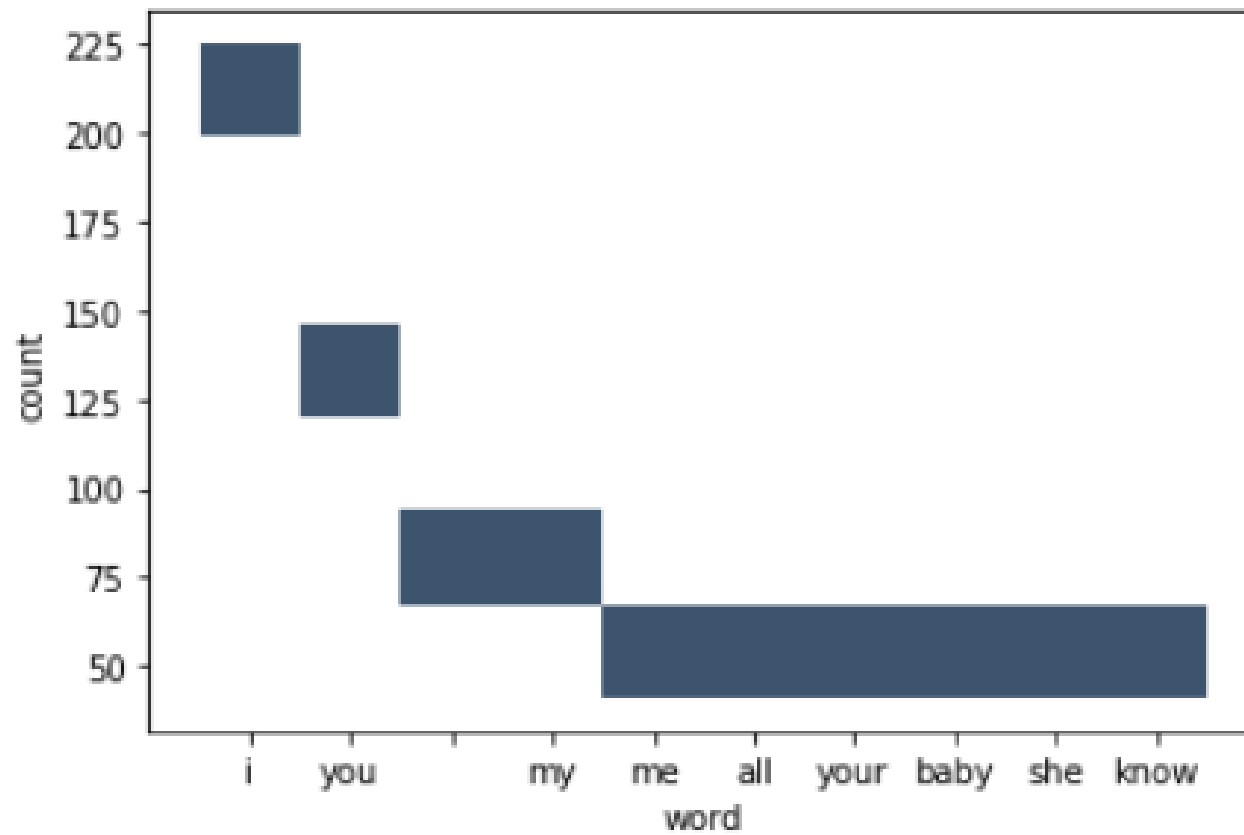
By looking at the most popular words, you will be able to find out the lyrics of songs which is kanye-styled:

Naive, self-confident, self-directed

	word	count
122	you	124
320	my	69
558	me	65
745	your	51
340	baby	48
901	she	42
823	know	41
638	but	41
180	lights	39
736	that	38

# WordCounting

---



Thanks for your  
attention!

---