enginius
MARKETING ENGINEERING ONLINE

Enginius

Sentiment Analysis

Sri Harsha Kommineni, University of Tampa

Table of Contents

Sentiment options

Options selected

Option	Selection
Data source	Sentiment data
Verbatim	Verbatim
Include date	Yes
Include rating	Yes
Word co-occurrence analysis and RAKE	Yes
Topic model	Yes
Number of topics	5
Default stop words	Long list (851)
Custom stop words	Yes
Date and time	2024-11-29 00:13:32 UTC

Options selected.

Data description

	Data	Number of Rows	Number of columns	Column names
1	Sentiment data	74	4	S.No, Verbatim, Date, Rating
2	Custom stop words	8	2	\, Stop words

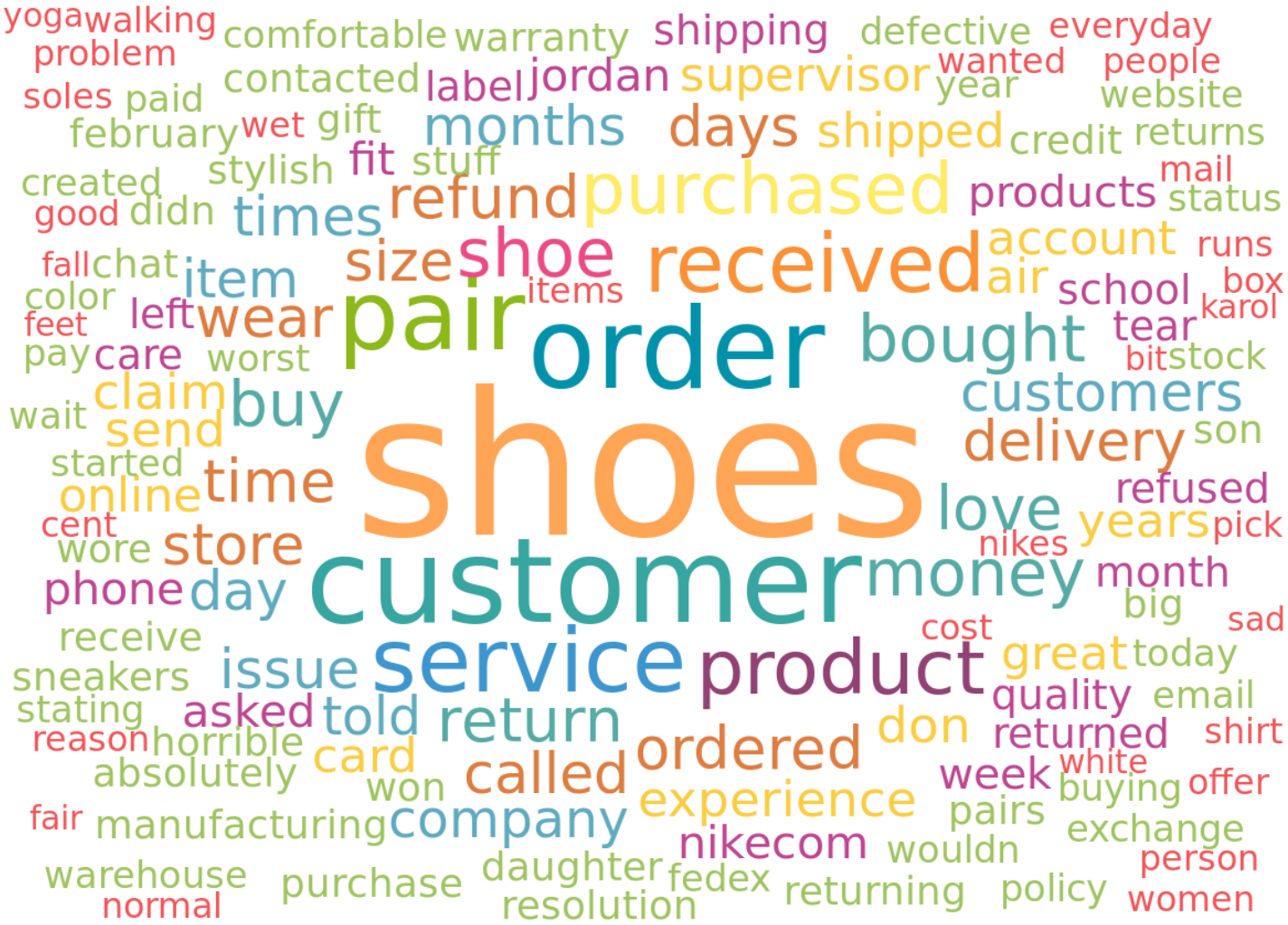
Data description.

Word cloud

Word cloud of the most frequent words

Word cloud. The word cloud represents the most frequently used words inside the corpus of texts provided. The bigger a word appears, the larger the number of times it occurs in the text corpus.

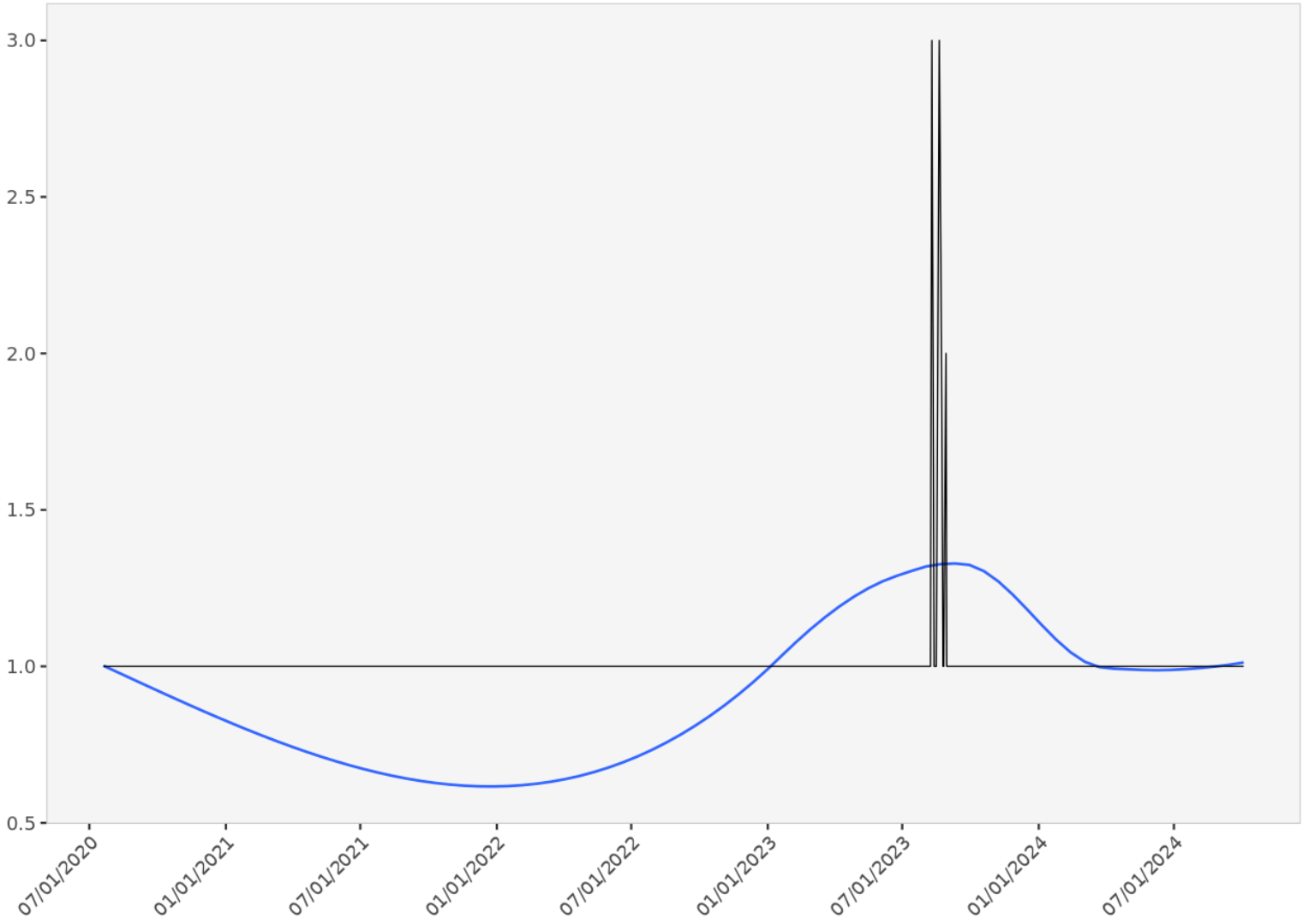
Word cloud of the most frequent words (without stemming)



Word cloud without stemming. The word cloud represents the most frequently used words inside the corpus of texts provided. The bigger a word appears, the larger the number of times it occurs in the text corpus.

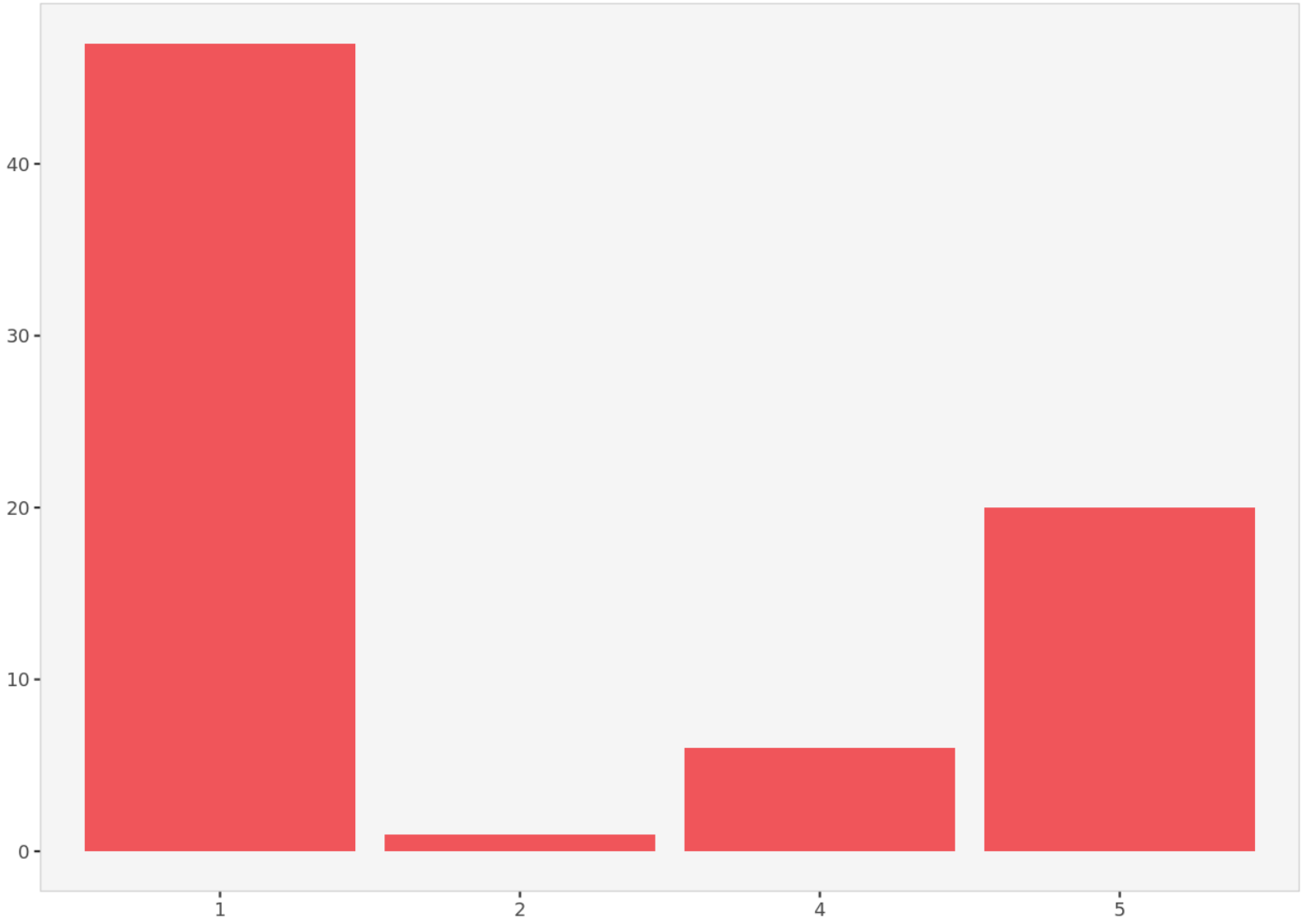
Sentiment analysis overview

Post frequency



Post frequency. The post frequency graph indicates the daily frequency of posts. The blue smoothing line helps visualize the trend.

Rating histogram



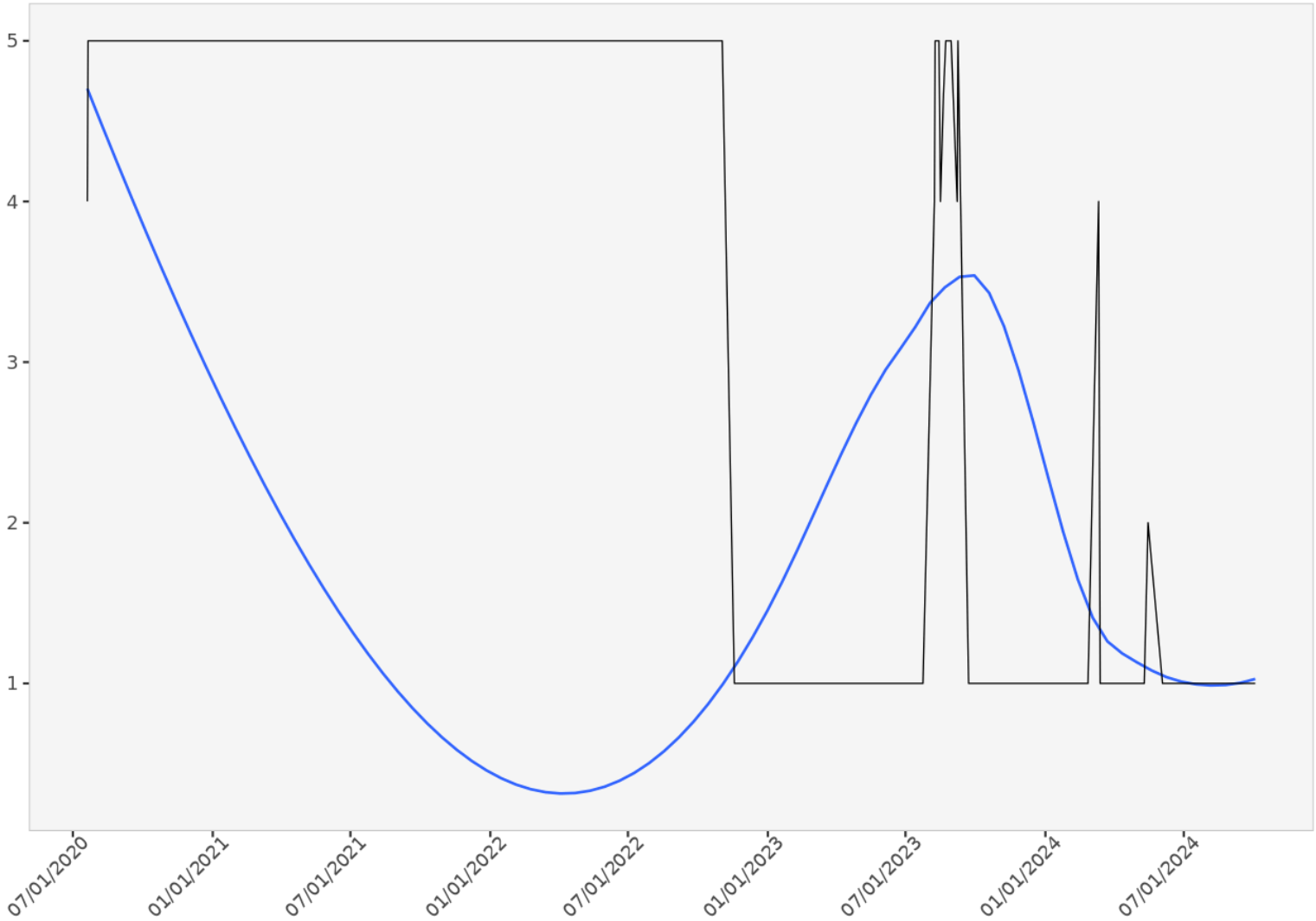
Rating histogram. The rating histogram indicates the number of posts categorized by their valence.

Rating frequency

	Frequency	Relative frequency
Total	74	100%
1	47	64%
2	1	1%
4	6	8%
5	20	27%

Rating frequency. Row names indicates the rating value, then frequency gives the number of post associated to this rating.

Average rating by dates



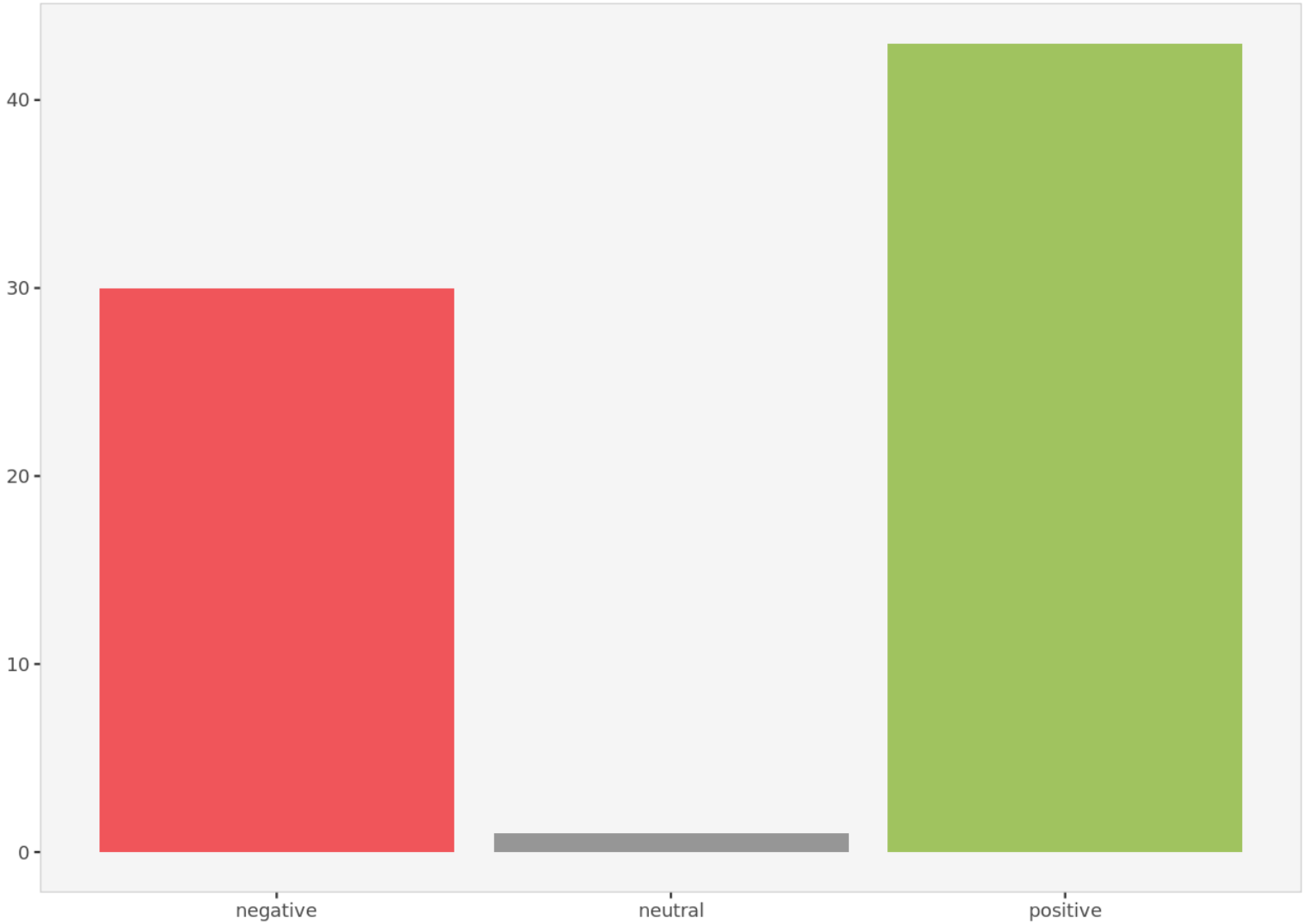
Average rating. The average rating graph indicates for each date the average rating of the posts. The blue smoothing line helps to visualize the trend.

Valence analysis

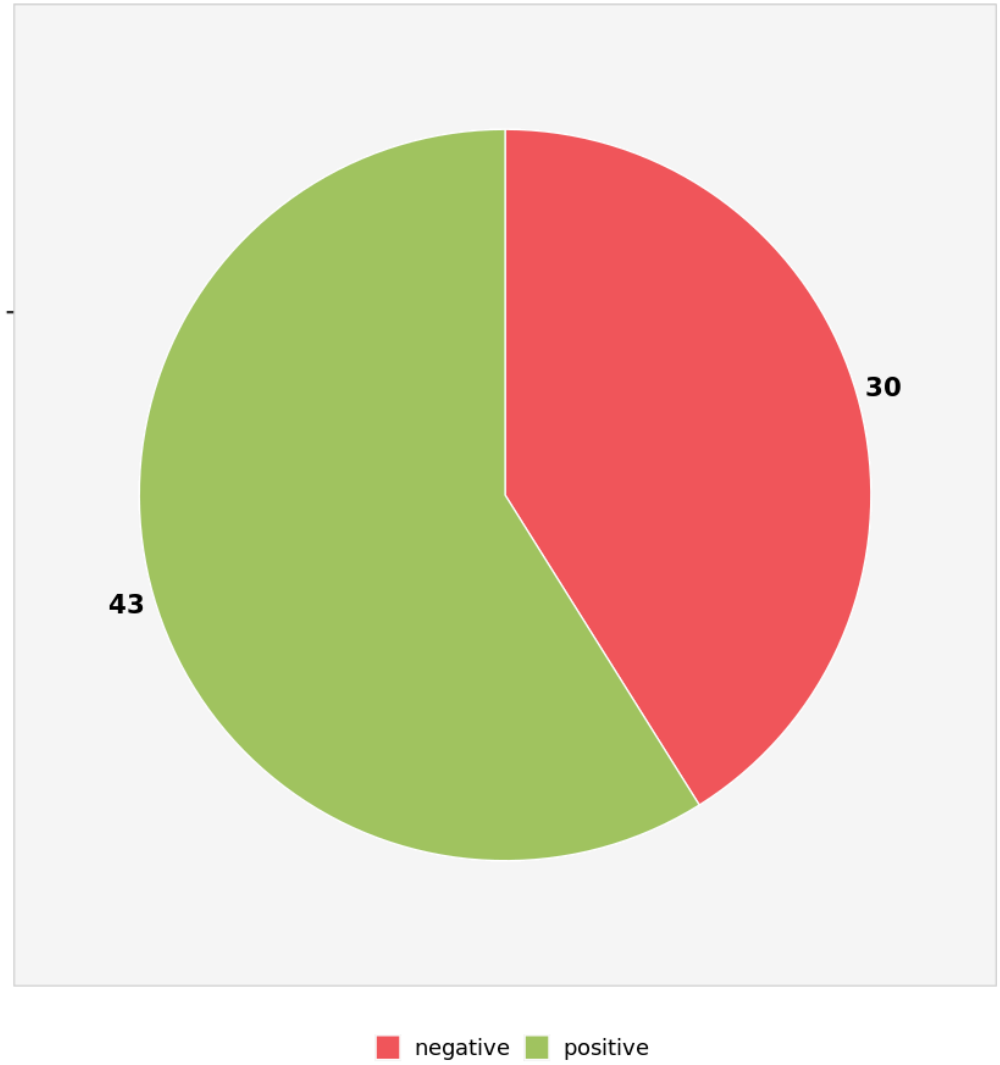
Valence repartition

	Posts count	Relative posts count
Total	74	100%
negative	30	41%
neutral	1	1%
positive	43	58%

Valence repartition. The number of posts that fall into different valence categories summarized by their absolute and relative values.

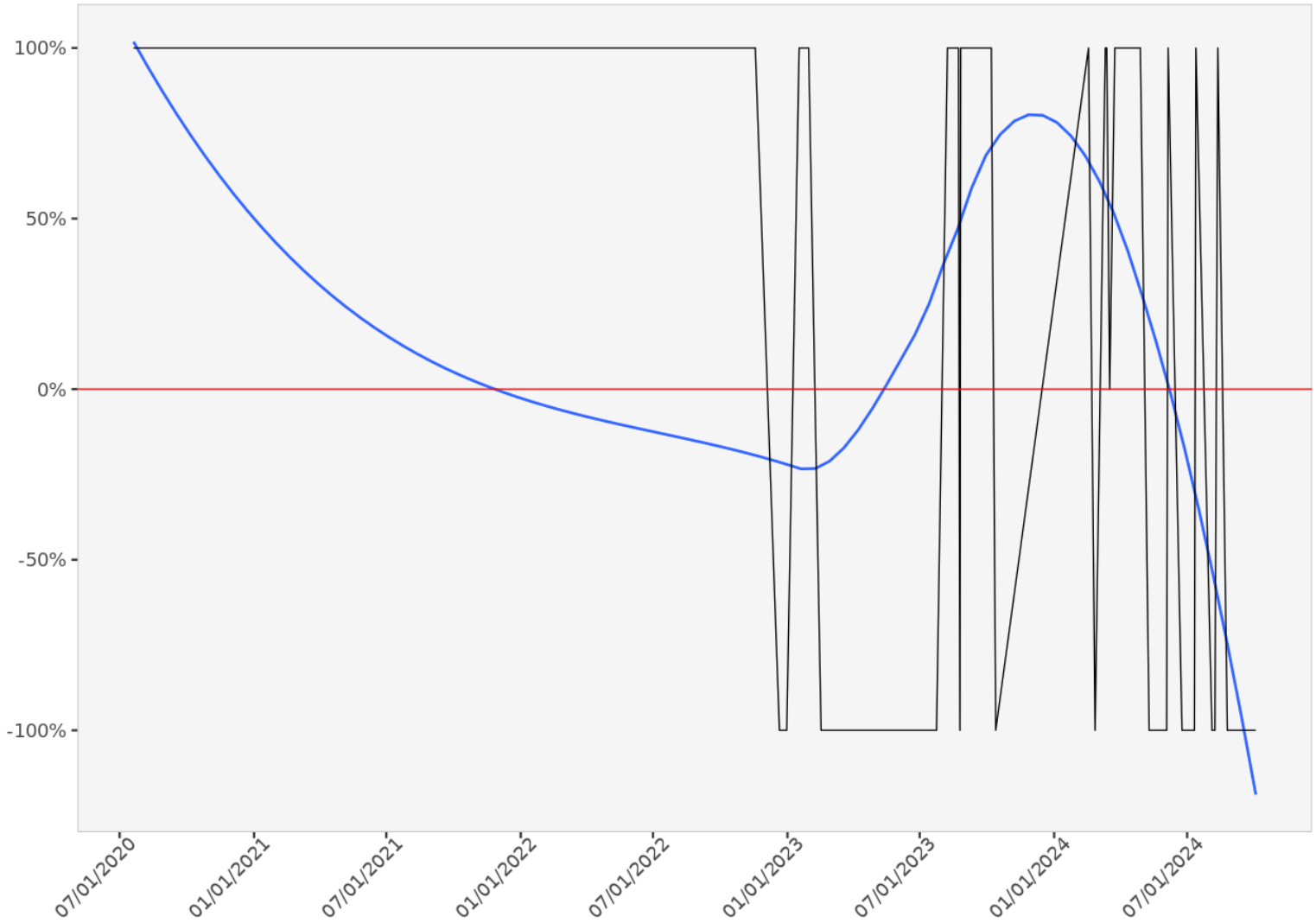


Valence histogram. The valence histogram indicates the number of posts by their valence.



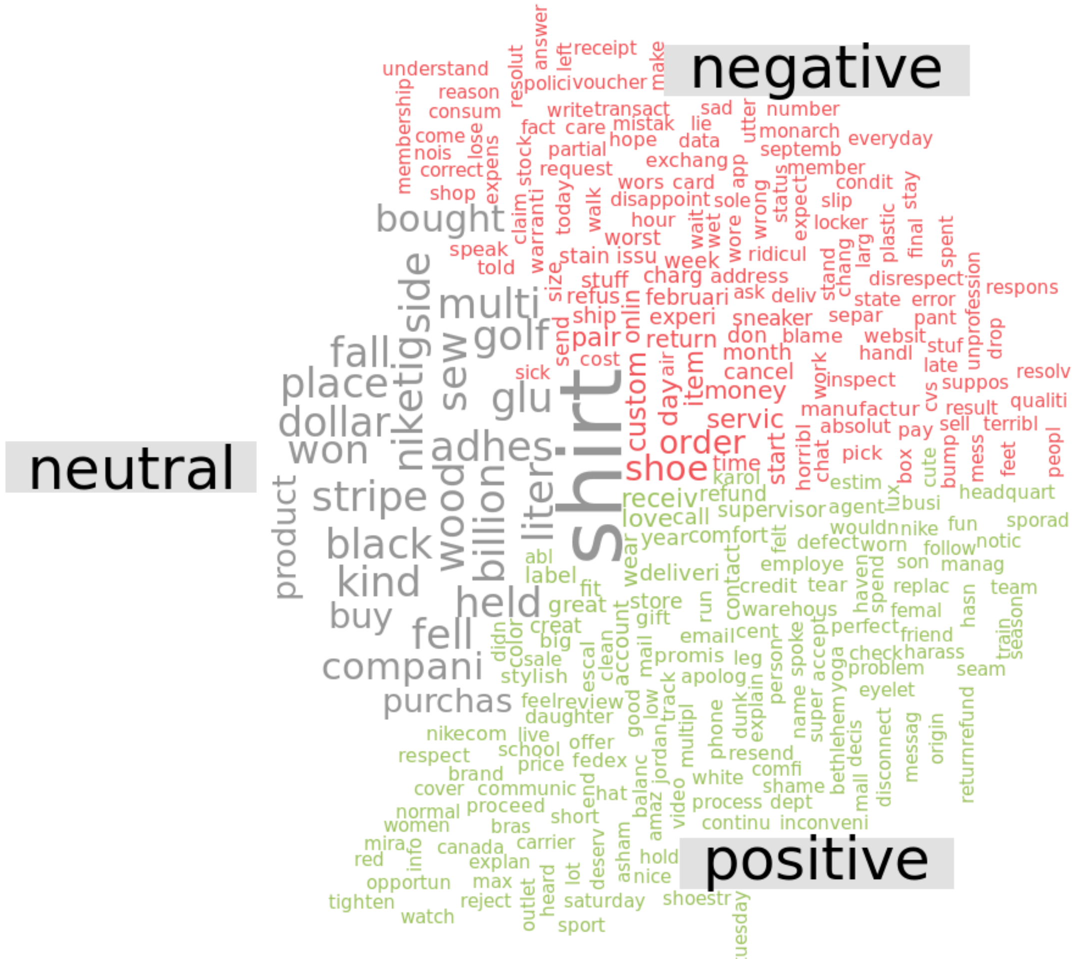
Valence distribution without uncategorized posts. Graphic summary of the relative sizes of the number of posts classified by valence after ignoring posts that could not be categorized (i.e., neutral posts).

Valence evolution



Post valence ratio. The post valence ratio graph indicates the daily average number of positive posts. The blue smoothing line helps visualize the trend.

Valence word cloud



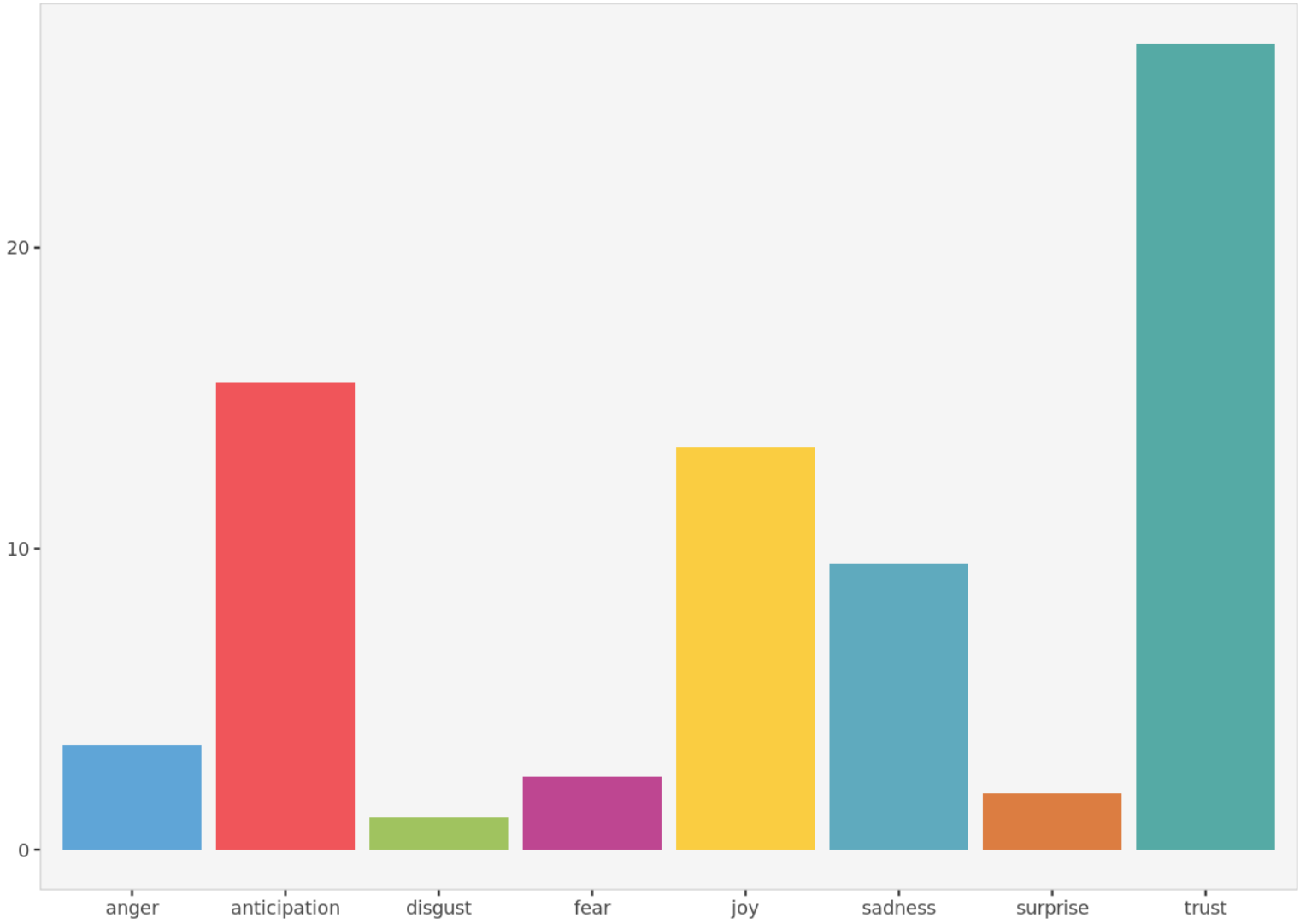
Valence word cloud.

Emotion analysis

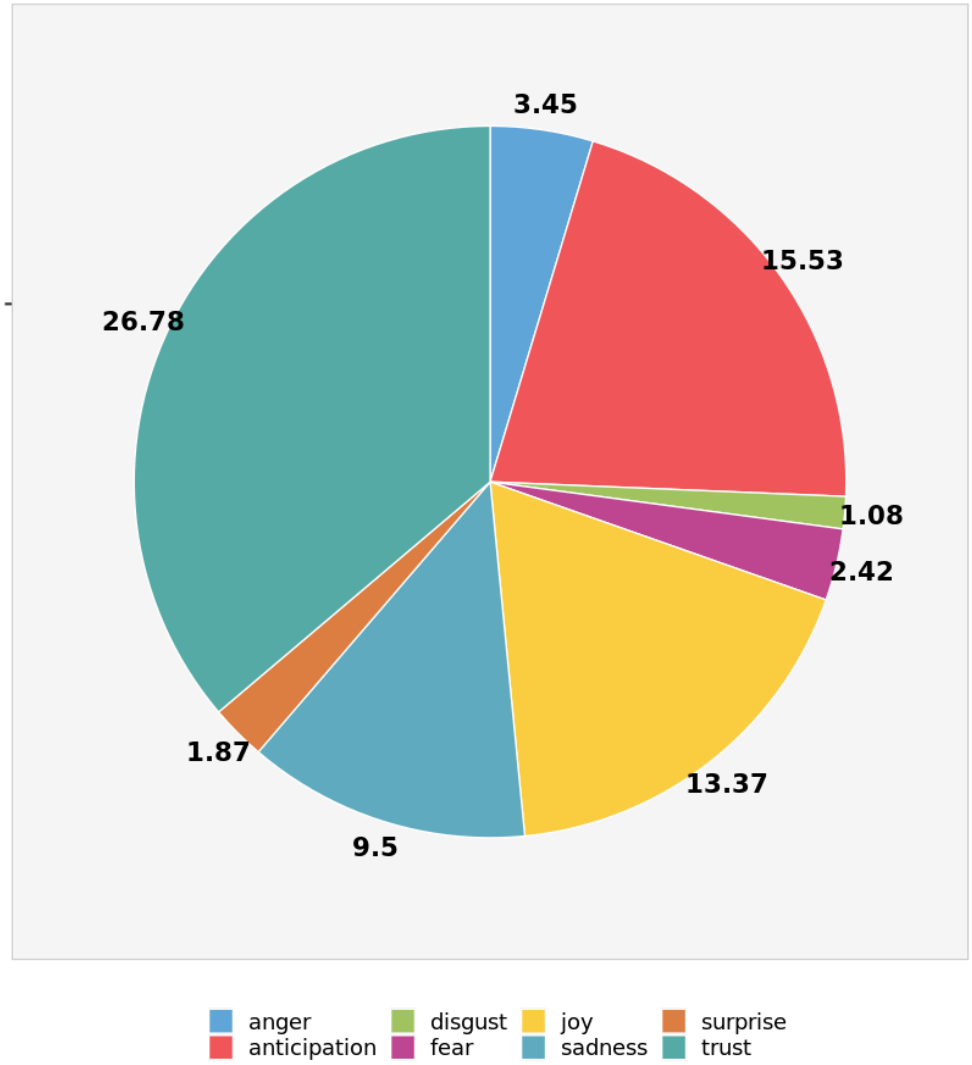
Emotion repartition

	Posts count	Relative posts count
Total	74.00	100%
anger	3.45	5%
anticipation	15.53	21%
disgust	1.08	1%
fear	2.42	3%
joy	13.37	18%
sadness	9.50	13%
surprise	1.87	3%
trust	26.78	36%

Emotion repartition. The number of posts that fall into different emotion categories summarized by their absolute and relative values. If a post has multiple emotions, then it is equally divided among those emotions.

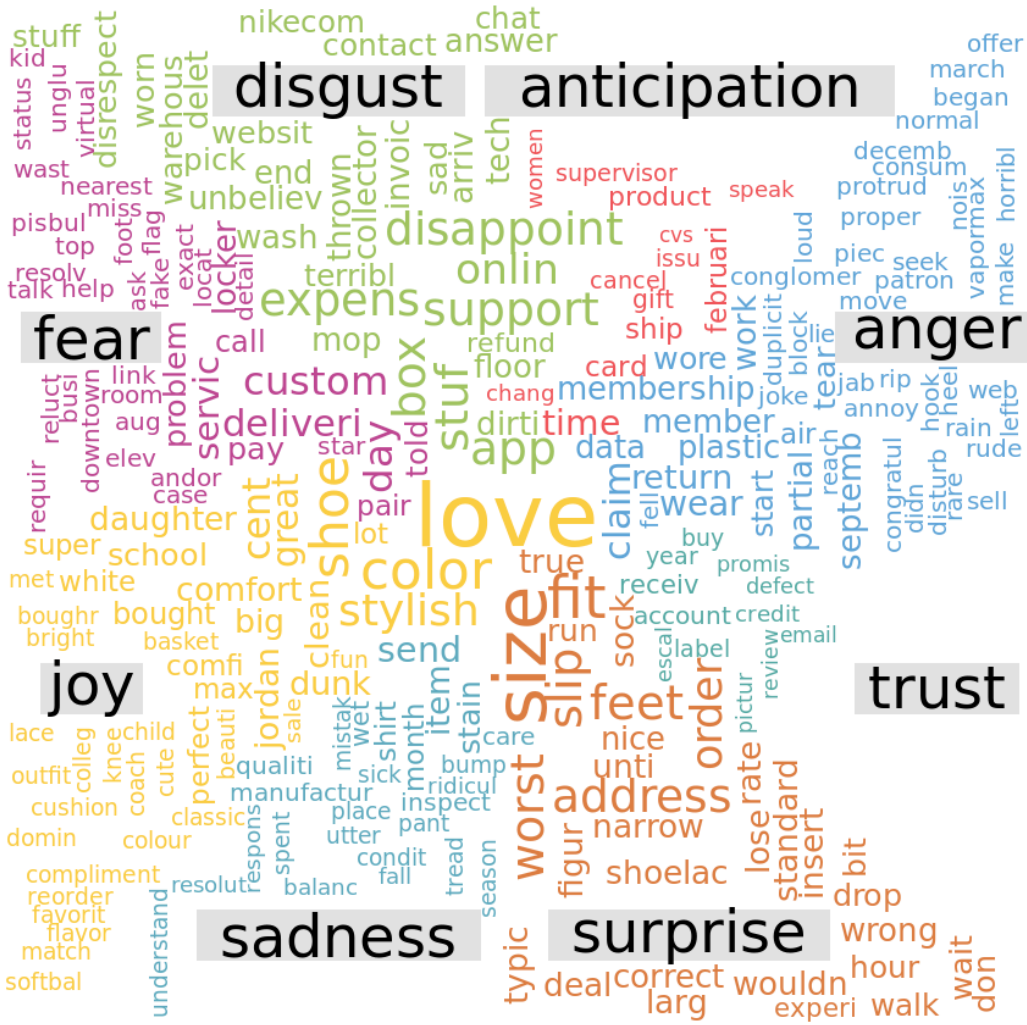


Emotion histogram. The emotion histogram indicates the number of posts by their emotion.



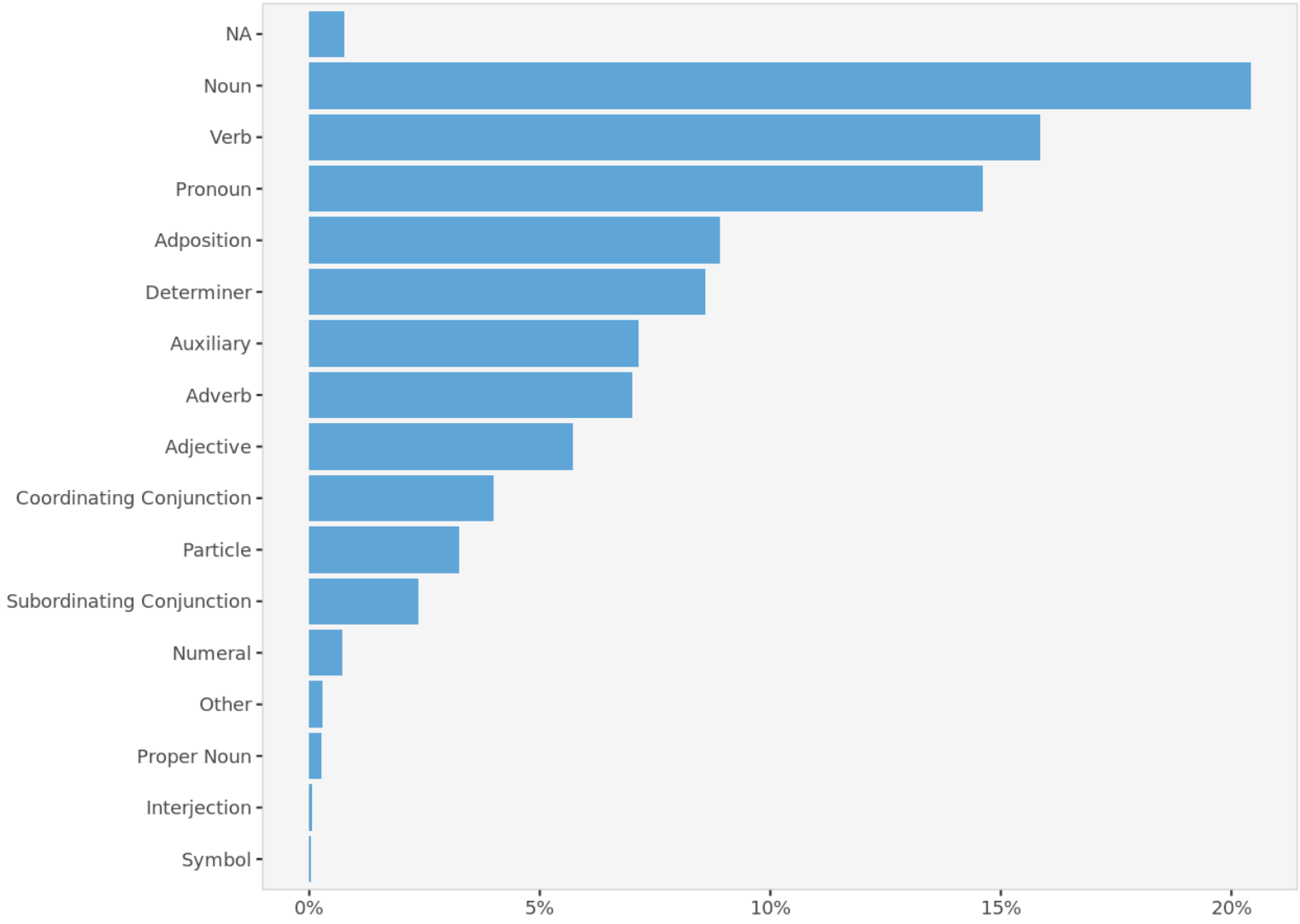
Emotion distribution without uncategorized posts. Graphic summary of the relative sizes of the number of posts classified by emotion after ignoring uncategorized posts.

Emotion word cloud



Emotion word cloud. Even if a post has multiple emotions its words will be shown only in one of those emotions.

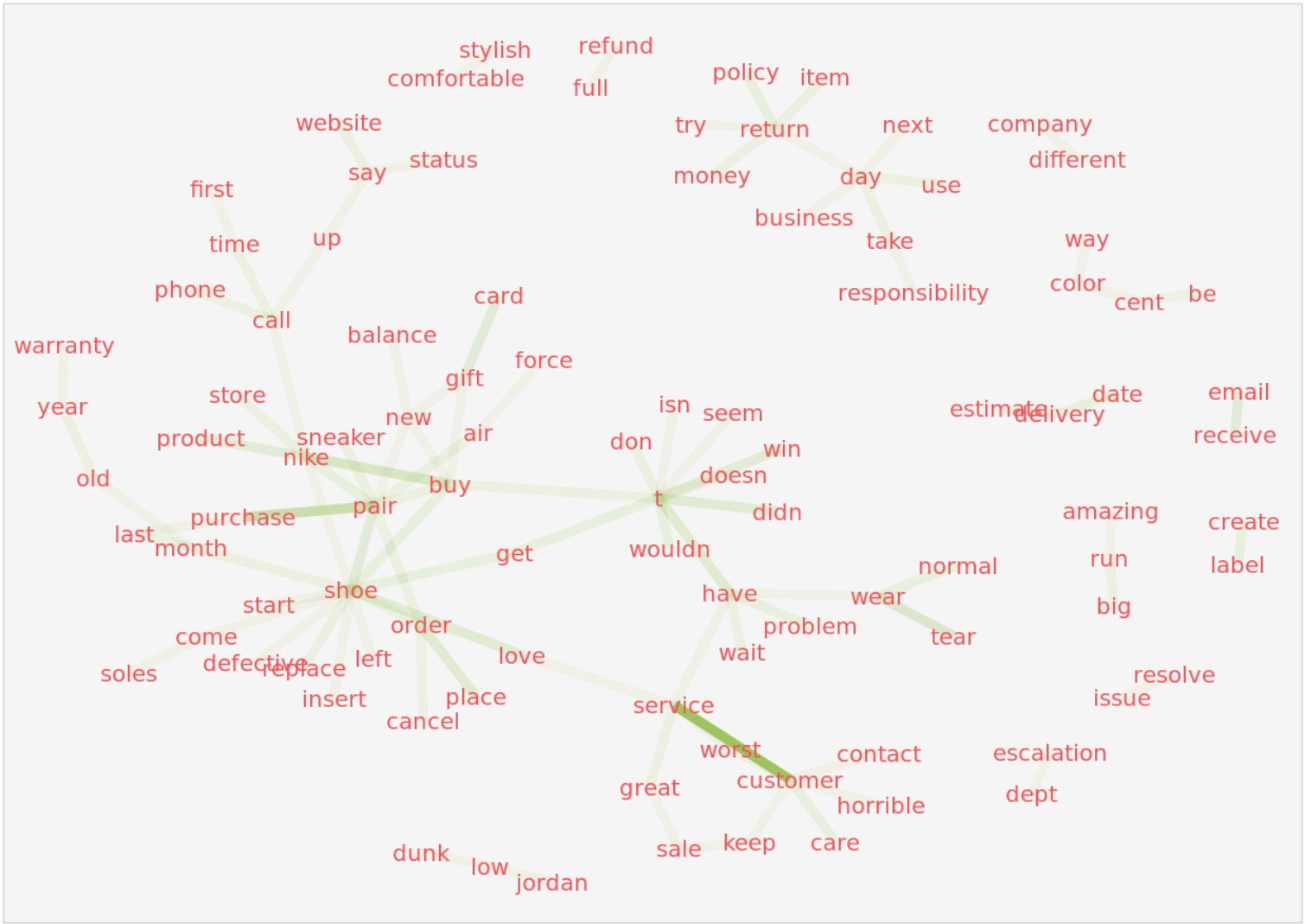
Distribution of Universal Parts of Speech Tags



Distribution of Universal Parts of Speech Tags.

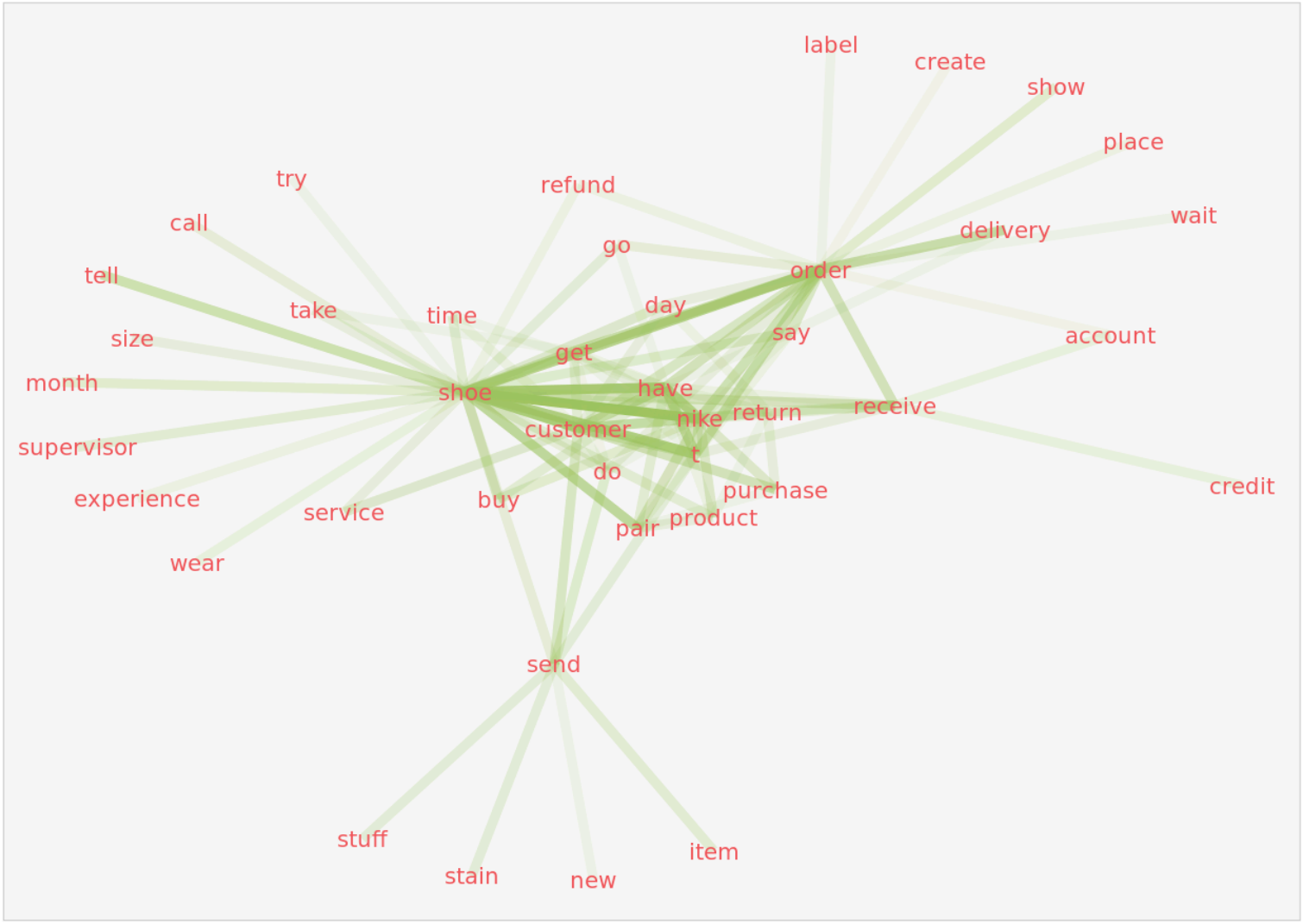
Graph of Word Co-Occurrences

Word co-occurrences between adjacent words in a corpus



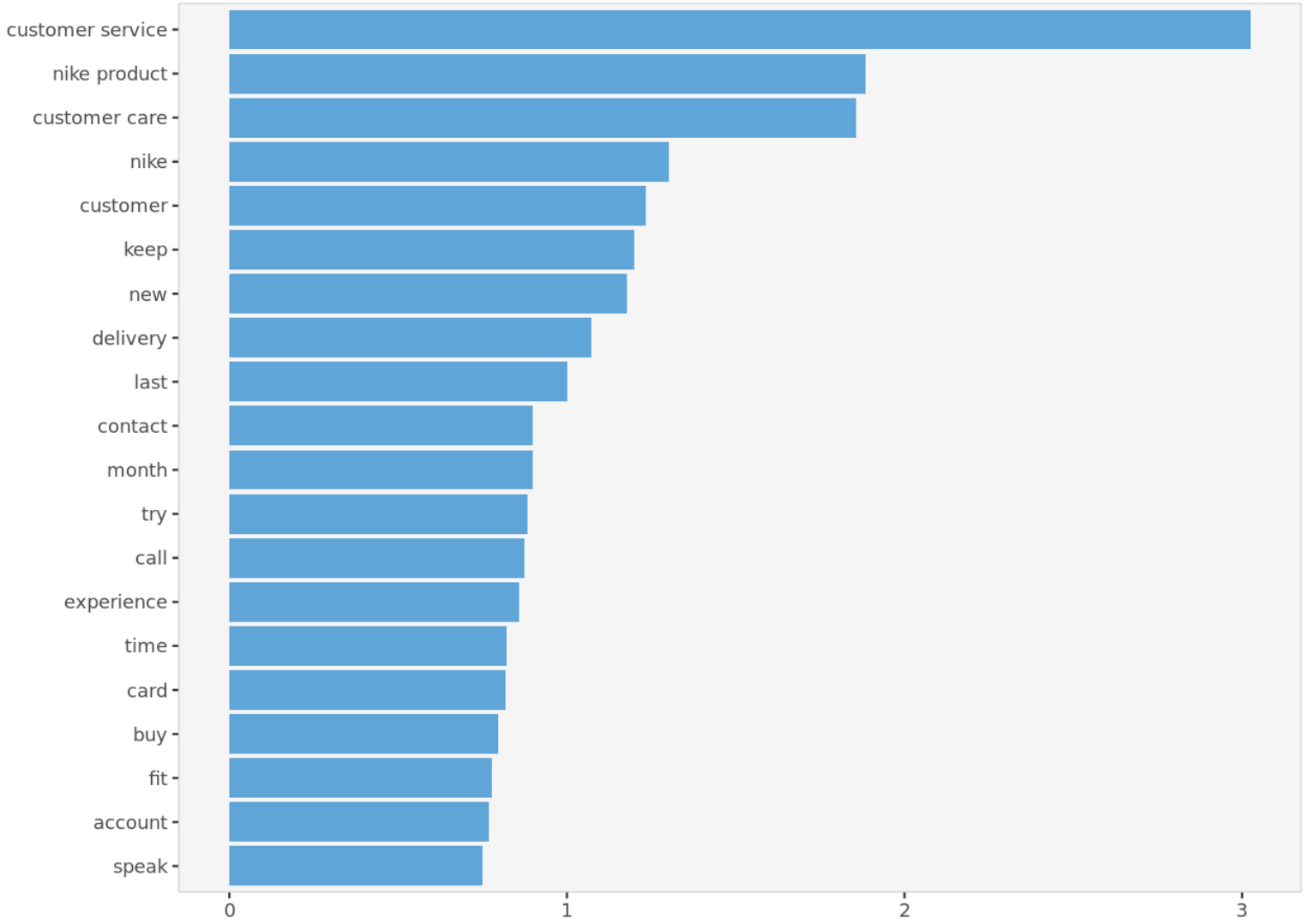
Includes Nouns, Verbs and Adjectives (retains custom stop-words).

Word co-occurrences within documents in a corpus



Includes Nouns, Verbs and Adjectives (retains custom stop-words).

RAKE Analysis (Rapid Automatic Keyword Extraction) -- Keywords with highest RAKE values



Keywords with highest RAKE values. The top keywords (i.e., contiguous sequence of words ignoring irrelevant words) were identified with minimum frequency of occurrences set to 0.01% of total word count.

Topic Model

Topics

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	receiv	love	shoe	pair	custom
2	order	wear	day	return	servic
3	time	year	told	purchas	order
4	ship	claim	experi	product	money
5	deliveri	month	care	buy	item
6	size	don	ask	store	send
7	wait	great	supervisor	bought	compani
8	week	manufactur	month	refund	account
9	nikecom	jordan	defect	card	onlin
10	refus	color	qualiti	air	call
11	shoe	sneaker	chat	state	school
12	call	comfort	issu	start	exchang
13	issu	bought	warranti	contact	credit
14	label	tear	replac	charg	big
15	run	fit	son	gift	absolut

Summary of top 15 words (ordered according to their importance) for the top 5 topics in the text corpus (also ordered according to their importance) in the text corpus (ignores custom stop-words).

Mean topic coherence = -221.376

Coherence is a metric based on the co-occurrence of the top words (say top 15) within a topic in each document of the corpus. For each pair of words in a topic, we compute the log of the probability that a document containing the higher ranked word also contains at least one instance of the lower-ranked word. The overall coherence value for a topic is the sum of scores for each pair of words. A number close to 0 (the highest possible value of this metric) indicates high coherence. We report the average coherence value across the topics.

Mean topic exclusivity = 14.555

Exclusivity, measured using a metric called FREX, captures the extent to which the top words in a topic are exclusive to that topic (i.e., are not as likely to occur in the other topics). The exclusivity score for each top word in a topic is the harmonic mean of two equally-weighted components: (1) the rate at which the word occurs within a topic relative to its rate of occurrence in the other topics, and (2) the frequency (number of times) of a word's occurrence within a topic relative to its frequency in the other topics. The exclusivity score for a topic is the average of the exclusivity scores of the words in a topic. The computed score is a positive number (> 0), with values substantially greater than 1 indicating a topic's high exclusivity. We report the average exclusivity value across the topics.