

Task 1

A twitter keyword search after google translation of “procurement” into “採購”, finds 204 results. Even with n = 10000 and trying both retryonratelimit = F and = T, we only see 204 results.

created_at	id	id_str	text	full_text
2023-03-18 10:22:35	1.637097e+18	1637097167969198081	上億預備金採購戰時傷藥 衛福部：依法盤點戰備醫藥物資...	上億預備金採購戰時傷藥 衛福部：依法盤點戰備醫藥物資...
2023-03-18 09:20:12	1.637081e+18	163708147103954305	上億預備金採購戰時傷藥 衛福部：依法盤點戰備醫藥物資...	上億預備金採購戰時傷藥 衛福部：依法盤點戰備醫藥物資...
2023-03-18 09:10:34	1.637079e+18	1637079044792680453	烏俄戰爭爆發造成大量傷亡，也有不少需要藥物治療的傷...	烏俄戰爭爆發造成大量傷亡，也有不少需要藥物治療的傷...
2023-03-18 09:09:57	1.637079e+18	163707888395448321	賴立雄：中共對台認知距離演變強烈 估大規模干預明年總統大...	賴立雄：中共對台認知距離演變強烈 估大規模干預明年總統大...
2023-03-18 08:45:45	1.637073e+18	163707280016060032	@cooksyz34 你真的有想要大搖擺? 如果有的話，我朋友在...	@cooksyz34 你真的有想要大搖擺? 如果有的話，我朋友在...
2023-03-18 07:34:42	1.637055e+18	1637054920586301440	@cwcty 不知道現在採購假睫毛 眼皮貼 是用自己的嗎?	@cwcty 不知道現在採購假睫毛 眼皮貼 是用自己的嗎?
2023-03-18 07:30:06	1.637054e+18	1637053761939210240	烏克蘭使用美國援助的陸射小直徑炸彈(GLSDB)，發揮了速...	烏克蘭使用美國援助的陸射小直徑炸彈(GLSDB)，發揮了速...
2023-03-18 06:31:16	1.637039e+18	1637038957501465056	戰備藥物採購 陳建仁：依法執行接受國會監督 - 中央通訊社...	戰備藥物採購 陳建仁：依法執行接受國會監督 - 中央通訊社...
2023-03-18 04:32:16	1.637009e+18	1637009008388366336	五艘「反艦型」沱江級艦採購案已在本月15日決標，由龍德...	五艘「反艦型」沱江級艦採購案已在本月15日決標，由龍德...
2023-03-18 04:11:47	1.637004e+18	1637003856008863744	完美的示範了達成推坑採購，我要封鎖對方5分鐘。https://t...	完美的示範了達成推坑採購，我要封鎖對方5分鐘。
2023-03-18 03:53:45	1.636999e+18	1636999316668813312	@EricChen0312 一家人食材採購一個月份量不到兩百元100万...	@EricChen0312 一家人食材採購一個月份量不到兩百元100万...
2023-03-18 02:53:08	1.636984e+18	1636984061335056386	每年的清明前夕，都要上山探望祖先。祖先曾傳達，希望有...	每年的清明前夕，都要上山探望祖先。祖先曾傳達，希望有...
2023-03-18 02:04:20	1.636972e+18	163697177984453890	BLACKPINK 不是晚上開場嗎，為什麼現在就有一批人從世運...	BLACKPINK 不是晚上開場嗎，為什麼現在就有一批人從世運...
2023-03-18 01:12:21	1.636959e+18	1636958699414781952	您只需要一鍵將它們上傳到您的商店。無需倉儲、無需接...	您只需要一鍵將它們上傳到您的商店。無需倉儲、無需接...
2023-03-18 00:11:59	1.636944e+18	1636943505259102211	下班了帶小孩採購 #mp100 #mobpsycho100 https://t.co/I...	下班了帶小孩採購 #mp100 #mobpsycho100 https://t.co/I...

After attempting to clean up the corpus, terms were sorted by frequency.

	freq	word
	9	✓
衛福部：依法盤點戰備醫藥物資	7	衛福部：依法盤點戰備醫藥物資
民視新聞網	6	民視新聞網
taiuanlang	4	taiuanlang
台東釋迦出頭天	3	台東釋迦出頭天
日通路商大手筆採購	3	日通路商大手筆採購
「	3	「
zedongwinine	3	zedongwinine
enlightenedev	3	enlightenedev
上億預備金採購戰時傷藥	3	上億預備金採購戰時傷藥
🔥俄烏戰爭爆發後導致國際情勢不穩，衛福部盤點國內戰備...	3	🔥俄烏戰爭爆發後導致國際情勢不穩，衛福部盤點國內戰備...
🔥上億預備金採購戰時傷藥	3	🔥上億預備金採購戰時傷藥
上億預備金採購戰時傷藥	3	上億預備金採購戰時傷藥
日本引進戰斧飛彈應對威脅！	2	日本引進戰斧飛彈應對威脅！
lelechen	2	lelechen
甲曲威桂竹青	2	甲曲威桂竹青

Showing 1 to 16 of 855 entries, 2 total columns

Due to how terms work in the language, splitting is a little difficult. The tools/libraries might be expecting spacesHowever, a few junk terms are blacklisted:

```
blacklist = c('✓', 'taiuanlang', '「', 'zedongwinine', 'enlightenedev', 'lelechen')
```

Check emoji, a few username handles, and the square open quotation mark 「 are blacklisted and removed.

```
95  
96 # get rid of unwanted words # this is changed too for the twitter analysis  
97 blacklist = c('✓','taiuanlang',' ','zedongwinine','enlightenedev','telechen')  
98 # d redefined to be not those rows whose word bit is in the blacklist  
99 # %in% is special R operator and returns stuff in a which are in b. (a %in% b)  
100 d = d[!d$word %in% blacklist, ]  
101  
102 #
```

A wordcloud is generated and saved, limited to terms which appear more than 3 times (min frequency of 3. While doing so, RStudio gave errors and skipped terms unable to fit on the image.



I can't actually read Chinese, so I put the terms/phrases through google translate. These are probably full sentences or verses. The documenttermmatrix() used to segment out terms is focused on English stop words, space as word separators, etc.

CHINESE (TRADITIONAL) - DETECTED	ENGLISH	CHINESE (TRADITIONAL)	ENGLISH	SPANISH
衛福部：依法盤點戰備醫藥物資	X	Ministry of Health and Welfare: Inventory of medical supplies for combat readiness according to law		☆
民視新聞網		FTV News Network		
台東釋迦出頭天		Taitung Sakyamuni first day		
日通路商大手筆採購		Nitto dealers make big purchases		
上億預備金採購戰時傷藥		Hundreds of millions of reserve funds to purchase wartime wound medicine		
🔥俄烏戰爭爆發後導致國際情勢不穩，衛福部盤點國內戰備藥物庫存，並申請第二預備金採購「外傷用戰略物資」以為因應。		🔥The outbreak of the Russo-Ukraine War led to instability in the international situation. The Ministry of Health and Welfare took stock of domestic war preparation medicines and applied for the second reserve fund to purchase "strategic materials for trauma" in response.		
🔥🔥上億預備金採購戰時傷藥		🔥🔥 Hundreds of millions of reserve funds to purchase wartime wound medicine		
上億預備金採購戰時傷藥				

There are issues with the small corpus (~200, rather than expected 10000 results), and complications of cleaning/splitting up another language's words. Then common “terms” in the first 200 of procurement seem to float around military procurement. It’s probably responses about a handful of news stories.

A search on twitter website for “採購”, is mostly military procurement news stories. It might be a term more commonly used for military procurement, or I might’ve searched at an unlucky time for procurement talk.

Task 1A – Poorly Implementing JieBaR to segment Chinese

This uses JieBa R library, which seems common for segmenting Chinese. I badly implemented based on lecture/blogpost found online: alvin chen workshop ¹, and a towardsdatascience blog post. ² While the blog post links a stopword list to use in the JieBaR worker, it wasn't used.

A search for "采购" on twitter is conducted, returning 357 results. Trying simplified to see if it's a different theme of tweets.

```
105 #### TASK 1A - Word cloud Procurement - bonus round: using jiebaR ####  
106 # STEP 1.1 - Extract from twitter ####  
107  
108  
109 # fetch via key-words  
110 procurement = search_tweets(q="采购",  
111     n=10000,  
112     include_rts=FALSE,  
113     lang="zh",  
114     retryonratelimit = F,  
115     token=atoken)  
116
```

	created_at	id	id_str	text
1	2023-03-28 05:57:00	1.640654e+18	1640654211615473665	美国和日本就清洁能源技术中用到的关键矿产达成贸易协议...
2	2023-03-31 01:18:43	1.641671e+18	1641671341664538624	@China_DongDi @Arsene74085883 @jcc143460098 @wanw...
3	2023-03-31 01:14:22	1.641670e+18	1641670247160901633	中共拉拢盟友推人民币结算，欲摆脱对美元依赖 中国推动...

Showing 1 to 4 of 357 entries, 43 total columns

Resulting pile is turned into a corpus as before, but first a JieBaR worker object is set up and used to segment() up the tweet fulltexts into many terms. This is further processed by term-documentmatrix() to remove punctuation marks and lowercase the latin alphabet terms. Stopwords might not apply, since the list is probably in English. Next step will further remove common but low-meaning words.

```
127 # using jiebaR library,  
128 # this follows alvin chen workshop : https://alvinntnu.github.io/presentations/Workshop-HKSU/docs/chinese-word-segmentation.html  
129  
130 install.packages("jiebaR")  
131 library("jiebaR")  
132 jbworker = worker()  
133  
134 # read in text  
135 # segment here is from jiebaR function to segment phrases into terms  
136 mycorpus = Corpus(VectorSource(segment(procurement$text, jbworker)))  
137  
138  
139 # cleanup the document  
140 # let's remove the 'stop words', mostly articles and heading numbers: 3.1, and, the, or  
141 # then we just have library full of key words  
142 # some of this might not work with chinese text (endings aren't really a thing, no lower/upper distinction)  
143 corpustdm = TermDocumentMatrix(mycorpus,  
144                                     control = list(removePunctuation = TRUE,  
145                                         stopwords = T,  
146                                         tolower = T,  
147                                         stemming = FALSE, # removes endings (-ing, -ed, etc),  
148                                         removeNumbers = T  
149                                     ))
```

¹ <https://alvinntnu.github.io/presentations/Workshop-HKSU/docs/chinese-word-segmentation.html>

² <https://towardsdatascience.com/chinese-natural-language-pre-processing-an-introduction-995d16c2705f>

Processing is same as before, blacklist differs:

```
163 # get rid of unwanted words # this is changed too for the twitter analysis  
164 blacklist = c('的','是','了','在','我','和','都','有','https','采购')  
165  
166
```

Without stop word removal, a lot more low-meaning terms hung around. This includes “possessive particle”, “is”, “le”, “has”, “is”. Due to links and the original search term, “https” and “procurement” are blacklisted.

A wordcloud is generated and saved, with minimum frequency of 20. I tried saving as a svg, but text did not draw properly (left). Might be a font-face issue. A PNG was saved (right) and cropped in word.



As with part 1, I still can't read Chinese. A lot of low-value words hanging around, but we never fed JieBaR Worker a stop-word dictionary. From further reading, cran's jiebar quick-start guide says jiebar worker has a keyword extractor mode. Maybe that function can be used in future work to pull out the high-meaning words.³

Business-wise, the big news story around time of extraction was a Yuan-denominated transaction for LNG for France via UAE from China.⁴ It's a significant move as oil/gas transactions are usually settled in US Dollar. A common term/keyword is the petrodollar, though I'm not seeing its equivalent in the cloud. It's one of the reasons for countries to hold US dollar reserves. This might be part of a longer trend in what currencies international trade is done in. The conflict in Ukraine and international economic maneuverings around it may be hastening the shift.

³ https://cran.r-project.org/web/packages/jiebaR/vignettes/Quick_Start_Guide.html

⁴ <https://www.nasdaq.com/articles/china-completes-first-yuan-settled-lng-trade>

Task 2 – word-cloud, sentiment analysis for: DHL-US (@DHLUS)

A twitter scrape for tweets on @DHLUS timeline results in 3250 tweets:

```
117 #### TASK 2 - @DHLUS wordcloud & sentiment analysis ####
118 #### STEP 2.1 - Extract from twitter ####
119
120 # fetch via username
121 dhlus_tweets = get_timeline(user="DHLUS", n=10000, token=atoken)
122 #### STEP 2.2 - WORD_CLOUD ####
```

	created_at	id	id_str	text
1	2023-03-09 11:13:56	1.633864e+18	1633863700519198721	RT @TeamKalitta: Following two da
2	2023-03-07 10:33:28	1.633129e+18	1633128740673798150	How much of your operating budg
3	2023-03-06 09:03:47	1.632744e+18	1632743781958144000	Working with a partner that priorit
4	2023-03-05 12:21:27	1.632431e+18	1632431139750027264	Which of the following is your busi
5	2023-03-05 11:11:04	1.632413e+18	1632413428873666562	RT @IndyCaronNBC: RETWEET to c
6	2023-03-04 12:42:42	1.632074e+18	1632074099009761282	@ex_oh_em Please send a DM to c

Showing 1 to 6 of 3,250 entries, 43 total columns

After cleaning by removal of stopwords, punctuation, forcing to lowercase, etc. the terms are sorted in descending order by frequency:

	freq	word
business	647	business
ecommerce	643	ecommerce
learn	547	learn
global	413	global
help	411	help
trade	333	trade
shipping	278	shipping
goglobal	246	goglobal
find	237	find
international	234	international
teamkalitta	232	teamkalitta
blog	224	blog
take	220	take
will	216	will
get	206	get
new	196	new
time	194	time
today	189	today
gogreen	171	gogreen
tips	163	tips

Blacklist to further clean up include own company name: DHL, DHLUS, and slogan “DHLDelivers”. Also includes “amp” as I assume it’s an artifact of ampersand sign: &.

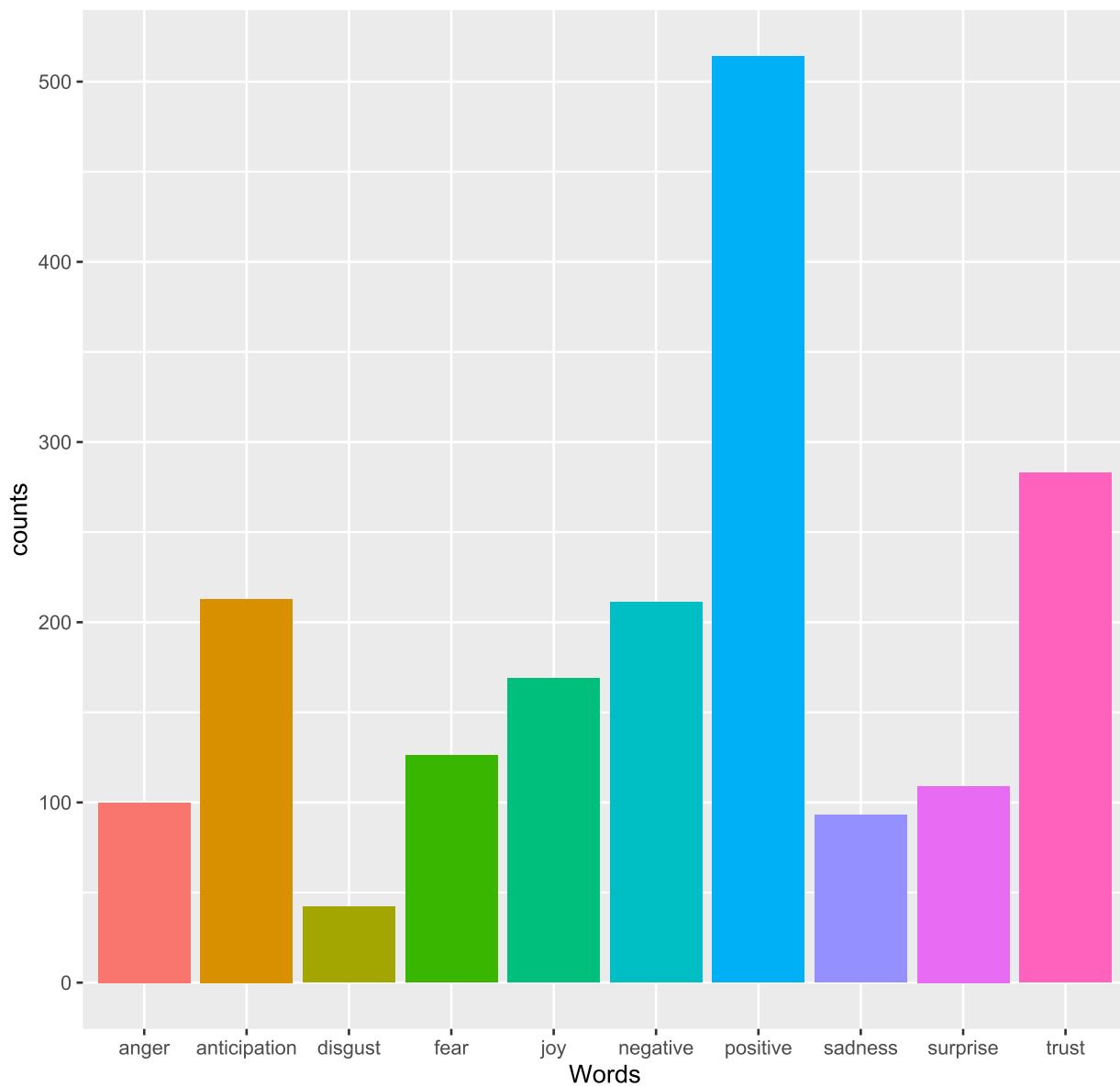
	id	id_str	text
1:00:09	1.321104e+18	1321104404544581633	#CybersecurityAwarenessMonth is here, but your security should
6:03:47	1.317557e+18	1317556937593024518	#CybersecurityAwarenessMonth is here. Get the facts your #busin
6:00:10	1.371189e+18	1371189432146128899	#DaylightSavingsTime springs forward on Saturday! If you do bus
6:00:32	1.298350e+18	1298349563674165248	#DHL is a proud leader in international shipping & #GoGreen
8:00:21	1.289682e+18	1289682408082571264	#DHL is a proud leader in international shipping & #GoGreen
3:30:13	1.263160e+18	1263160109271666689	#DHL is a proud leader in international shipping & #GoGreen
2:00:08	1.284881e+18	1284880714693931008	#DHL is a proud leader in international shipping & #GoGreen
6:00:21	1.315382e+18	1315381746251173890	#DHL is a proud leader in international shipping & #GoGreen

After removal of blacklisted terms, a wordcloud was generated for terms w/ frequency ≥ 120 .



Covid appears a couple times. TeamKalitta seems to be a DHL-sponsored racing team, with JRTodd a driver. Ecommerce and Business appear a lot. There's also a big focus on international/global commerce and sustainability (gogreen, sustainability). DHLUS twitter account also puts a lot of time in its racing team, team Kalitta, with driver jrtodd (@JRTodd373) and former driver Michael Andretti (@FollowAndretti) being mentioned.

Procurement Trend DHLUS 2023-Mar-27



A sentiment analysis based on “nrc” dictionary (terms to sentiments) is performed on the cleaned-up list of terms used in the tweets (d2). The plot above shows that @DHLUS account’s tweets have mostly used terms associated with trust, positive, and anticipation. Terms associated with negative emotions, fear, and sometimes anger do appear, however as logistics company DHLUS may be notifying people of outages and delays.

TASK 3 – “Inventory Management” – word cloud and sentiment analysis

Similar task to Task 2, except searching for tweets with term “Inventory management”. A query returns 1263 tweets.

	created_at	id	id_str	text
1	2023-03-27 02:14:01	1.640236e+18	1640235706047164416	Why do fasteners need smart shelves
2	2023-03-27 02:02:50	1.640233e+18	1640232894017814528	Revolutionize the way you do business
3	2023-03-27 02:00:09	1.640232e+18	1640232219527585793	Exciting news! 🎉 Our inventory man
4	2023-03-27 01:52:40	1.640230e+18	1640230335748231168	Internet or PC based #Software for G
5	2023-03-27 01:46:01	1.640229e+18	1640228662342561793	@Ssenonyiderick For inventory and
6	2023-03-27 01:34:31	1.640226e+18	1640225766523691012	god my maplestory inventory manag
7	2023-03-27 01:09:18	1.640219e+18	1640219422408441859	ClinicNEXT is an integrated solution w

Showing 1 to 7 of 1,263 entries, 43 total columns

A similar cleaning (removing punctuation, stop words, forcing to lower-case, etc), and splitting into key words and frequencies, leaves us with 7205 keywords.

	freq	word
inventory	1437	inventory
management	1307	management
software	291	software
business	231	business
can	215	can
amp	136	amp
system	126	system
accounting	126	accounting
erp	111	erp
supply	109	supply
warehouse	102	warehouse
like	97	like
help	94	help
chain	94	chain
control	88	control
get	84	get
ecommerce	83	ecommerce
just	83	just
✓	82	✓

Showing 1 to 19 of 7,205 entries, 2 total columns

Blacklist is made to remove a few terms:

- Common words like “can”, “like”, “get”, “just”.
 - Artifacts (amp from ampersand (&))
 - Check emoji.
 - “inventory” and “management” will always show up, being the keywords we searched..

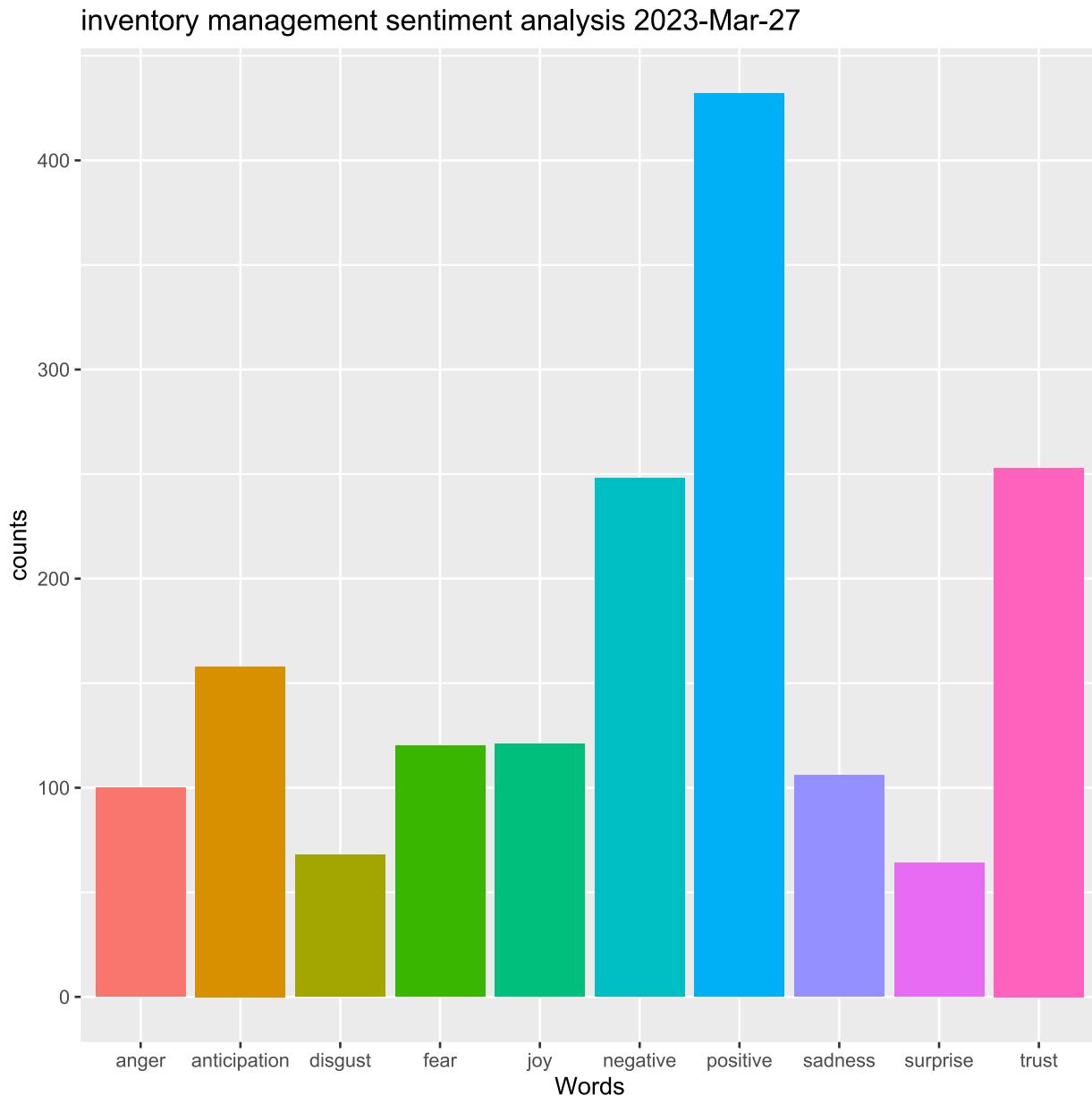
```
245  
246 # get rid of unwanted words # this is changed too for the twitter analysis  
247 blacklist = c('amp','like','just','can','get','inventory','management','will',  
248 # d redefined to be not those rows whose word bit is in the blacklist  
249 # %in% is special R operator and returns stuff in a which are in b. (a %in% b)  
250 d3 = d3[!d3$word %in% blacklist, ]
```

A wordcloud is generated from the resulting keywordlist: (image cropped for readability)



Looking at the word-cloud, a lot of talk about e-commerce, business-oriented, and software. Very common are accounting and enterprise-resource planning (software/systems). Overall, there is a lot of talk on software and computer systems to help account and plan supply/logistics. The issue of efficiency, control, and improvement/optimization show up.

Continuing onto the sentiment analysis of tweets featuring the term.



Words associated with inventory management are often associated with “positive” emotions. However, negative terms also appear in third, very close to the second place of trust-associated words. A lot of talk about inventory management may be to sell solutions, so positive and trust-worthy terms may be being used to promote inventory management solutions.

Negative, fearful, and anticipation associated terms may appear as well as inventory management is to help avoid risks of under- and over- stocking. Stocking too much will result in costs of depreciation, losses, and storage costs. Too little inventory runs risk of stock-out, and missed sales or project delays.