BAC 2025 Q2 Earnings Report Natural Language Processing in R

Isabel Anne Villa Mendoza

2025-08-22

Case Study Scope (Tech + Finance Integration)

Objective: Examine Bank of America's Q2 2025 earnings release using R programming and Natural Language Processing (NLP). Methodology: Applied sentiment analysis on the earnings call transcript and press release to quantify market tone. Integrated time-series modeling to assess how sentiment shifts impacted post-earnings stock price movements. Value-Add: Demonstrates ability to merge machine learning techniques with financial analysis, highlighting not only the earnings performance but also the behavioral market response.

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.2     ✔ tibble    3.3.0
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.1.0
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(textdata)
library(tidytext)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```r
library(RColorBrewer)
```

Load the Text File

```r
BACQ2 = read_file('BOFA_2025_Q2.txt')
```

```r
nchar(BACQ2)
```

```
## [1] 74486
```

```r
summary(BACQ2)
```

```
##    Length     Class      Mode
##         1 character character
```

```r
BACQ2 %>% stringr:: str_sub(1,999)
```

```
## [1] "Bank of America\nSecond Quarter 2025 Earnings Announcement\nJuly 16, 2025\n\n\fSecond Quarter 2025 Earnin
gs Announcement\nJuly 16, 2025\n\nParticipants\nPresenters\nBrian Moynihan – Bank of America, Chair and CEO\nAlas
tair Borthwick – Bank of America, CFO\nLee McEntire – Bank of America, Investor Relations & Local Markets Organiz
ation Executive\n\nParticipants\nBetsy Graseck - Morgan Stanley\nChris McGratty - Keefe, Bruyette, & Woods\nErika
Najarian - UBS\nGerard Cassidy - RBC Capital Markets\nJim Mitchell - Seaport Global\nJohn McDonald - Truist Secur
ities\nKen Usdin - Bernstein Autonomous\nMatt O'Connor - Deutsche Bank\nMike Mayo - Wells Fargo\nSteven Alexopoul
os - TD Cowen\n\nPresentation\nLee McEntire\nThank you, Chloe. Good morning, everyone. Thank you for joining us t
o review the second quarter results.\nOur earnings release documents are available on the Investor Relations sect
ion of the bankofamerica.com\nwebsite. Those documents include the earnings presentation that we'll make referenc
e to during the cal"
```

```r
BACQ2 %>% stringr:: str_count()
```

```
## [1] 74486
```

```r
#Look for keywords
BACQ2 %>% stringr::str_detect(c('positive', 'sales', 'growth', 'increase', 'volatility', 'expansion', 'improve'))
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```r
#Count the number of matches of a substring
BACQ2 %>% stringr::str_count(('positive'))
```

```
## [1] 4
```

```r
BACQ2 %>% stringr::str_count(('sales'))
```

```
## [1] 13
```

```r
BACQ2 %>% stringr::str_count(('increase'))
```

```
## [1] 14
```

```r
BACQ2 %>% stringr::str_count(('volatility'))
```

```
## [1] 3
```

```r
BACQ2 %>% stringr::str_count(('expansion'))
```

```
## [1] 3
```

```r
BACQ2 %>% stringr::str_count(('improve'))
```

```
## [1] 24
```

```r
#Where are the keywords mentioned?

BACQ2%>% stringr::str_locate_all('positive')
```

```
## [[1]]
##      start   end
## [1,] 19797 19804
## [2,] 43063 43070
## [3,] 72238 72245
## [4,] 72513 72520
```

```r
BACQ2%>% stringr::str_locate_all('sales')
```

```
## [[1]]
##       start   end
##  [1,]  4540  4544
##  [2,]  8256  8260
##  [3,]  8961  8965
##  [4,] 13525 13529
##  [5,] 22938 22942
##  [6,] 24653 24657
##  [7,] 25220 25224
##  [8,] 28749 28753
##  [9,] 31784 31788
## [10,] 31859 31863
## [11,] 36148 36152
## [12,] 67148 67152
## [13,] 67202 67206
```

```r
BACQ2%>% stringr::str_locate_all('increase')
```

```
## [[1]]
##       start   end
##  [1,]  5816  5823
##  [2,]  6826  6833
##  [3,]  8370  8377
##  [4,] 15362 15369
##  [5,] 18385 18392
##  [6,] 19968 19975
##  [7,] 22857 22864
##  [8,] 29467 29474
##  [9,] 30985 30992
## [10,] 32334 32341
## [11,] 57895 57902
## [12,] 59300 59307
## [13,] 64865 64872
## [14,] 65039 65046
```

```r
BACQ2%>% stringr::str_locate_all('volatility')
```

```
## [[1]]
##      start   end
## [1,] 21203 21212
## [2,] 32125 32134
## [3,] 68164 68173
```

```r
BACQ2%>% stringr::str_locate_all('exoansion')
```

```
## [[1]]
##      start end
```

```r
BACQ2%>% stringr::str_locate_all('improve')
```

```
## [[1]]
##       start   end
##  [1,]  4774  4780
##  [2,]  9925  9931
##  [3,] 11711 11717
##  [4,] 12373 12379
##  [5,] 13680 13686
##  [6,] 15006 15012
##  [7,] 17904 17910
##  [8,] 21729 21735
##  [9,] 22442 22448
## [10,] 22892 22898
## [11,] 23880 23886
## [12,] 23977 23983
## [13,] 26681 26687
## [14,] 26745 26751
## [15,] 27036 27042
## [16,] 27174 27180
## [17,] 27271 27277
## [18,] 27557 27563
## [19,] 31726 31732
## [20,] 31893 31899
## [21,] 32438 32444
## [22,] 35625 35631
## [23,] 47086 47092
## [24,] 56379 56385
```

```r
#view surrounding text ( requires regex)
BACQ2 %>% stringr:: str_extract_all(".{20}(positive).{20}")
```

```
## [[1]]
## [1] "ill be NII slightly positive. So, you got a litt"
```

```r
BACQ2 %>% stringr:: str_extract_all(".{20}(sales).{20}")
```

```
## [[1]]
## [1] "r of year-over-year sales and trading growth,"
## [2] "nd quarter level of sales and trading revenue"
## [3] "r's $5.4 billion of sales and trading revenue"
## [4] "gement fees, higher sales and trading revenue"
## [5] "nd their associated sales, as I noted earlier"
## [6] "f last year on good sales and trading results"
## [7] " lower. Focusing on sales and trading ex DVA,"
## [8] "owth, obviously, in sales and trading, up 15%"
## [9] " double the size of sales force by converting"
```

```r
BACQ2 %>% stringr:: str_extract_all(".{20}(increase).{20}")
```

```
## [[1]]
## [1] " same time, we also increased the capital return"
## [2] "ast 6 years. We saw increases in the average con"
## [3] " more than 5%. That increase reflects an aggrega"
## [4] "activity grew as we increased investment for the"
## [5] "franchise as we saw increased activity across Eu"
## [6] "x it, we'll have an increase coming at us in ano"
## [7] "housing investments increased to offset that win"
```

```r
BACQ2 %>% stringr:: str_extract_all(".{20}(volatility).{20}")
```

```
## [[1]]
## [1] "ting from the macro volatility. Our Equities group"
## [2] "o be careful of the volatility of consumer credit "
```

```r
BACQ2 %>% stringr:: str_extract_all(".{20}(expansion).{20}")
```

```
## [[1]]
## [1] "ve led to continued expansion in digitalization a"
## [2] "t to do its organic expansion, and that's what we"
## [3] "uing that push. The expansion markets we call the"
```

```r
BACQ2 %>% stringr:: str_extract_all(".{20}(improve).{20}")
```

```
## [[1]]
##  [1] "e quarterly results improved as each month of t"
##  [2] " models and process improvements to help us get"
##  [3] "p, AI tools to help improve our process around "
##  [4] "tended to help both improve the client experien"
##  [5] " was led by the NII improvement, it also includ"
##  [6] "he efficiency ratio improved more than 200 basi"
##  [7] "billion with market improvement and full year f"
##  [8] "mer net charge-offs improved linked quarter, fo"
##  [9] "ing ex DVA, revenue improved 15% year-over-year"
## [10] "as up 9% on revenue improvement and continued i"
## [11] "utlook implied some improvement in the second h"
## [12] " And so, what's the improved process. And then "
```

Tokenize the text

```r
#Change to a tibble (tidy dataframe)
token1 <-tibble(BACQ2)

#Tokenize
token1 <- token1 %>% tidytext::unnest_tokens(output=word, input=BACQ2, token='words', to_lower=TRUE)
```

```r
view(token1)
```

```r
#add order of words
token1 <- token1 %>% mutate(order = row_number())
```

```r
#count tokens
token1 %>% nrow()
```

```
## [1] 13044
```

```r
#first few words
token1 [1:20,]
```

```
## # A tibble: 20 × 2
##    word         order
##    <chr>        <int>
##  1 bank             1
##  2 of               2
##  3 america          3
##  4 second           4
##  5 quarter          5
##  6 2025             6
##  7 earnings         7
##  8 announcement     8
##  9 july             9
## 10 16              10
## 11 2025            11
## 12 second          12
## 13 quarter         13
## 14 2025            14
## 15 earnings        15
## 16 announcement    16
## 17 july            17
## 18 16              18
## 19 2025            19
## 20 participants    20
```

```r
#count the number of matches of a sub-string
token1 %>% dplyr::filter(word==str_sub('growth')) %>% count()
```

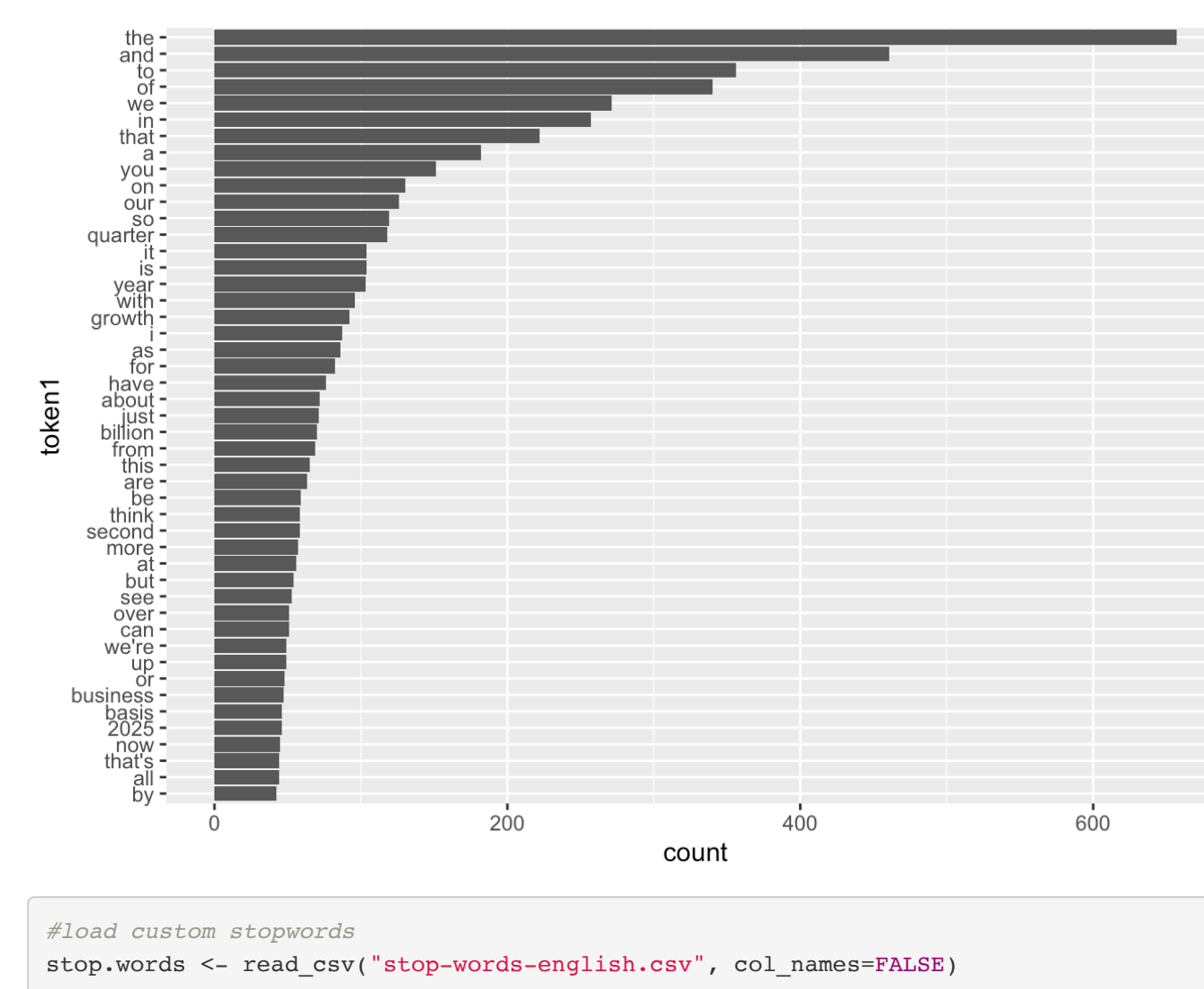
```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    92
```

```r
#where is the keyword mentioned
token1 %>% dplyr::filter(word==str_sub('growth'))
```

```
## # A tibble: 92 × 2
##    word   order
##    <chr>  <int>
##  1 growth   326
##  2 growth   330
##  3 growth   350
##  4 growth   483
##  5 growth   500
##  6 growth   672
##  7 growth   731
##  8 growth   924
##  9 growth  1010
## 10 growth  1374
## # ℹ 82 more rows
```

Remove Stop Words

```r
# Look at the most important frequent words
token1 %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 40) %>%
  mutate(token1 = reorder(word, count)) %>%
  ggplot(aes(x=count , y=token1)) +
  geom_col()
```



```r
#load custom stopwords
stop.words <- read_csv("stop-words-english.csv", col_names=FALSE)
```

```
## Rows: 1298 Columns: 1
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): X1
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

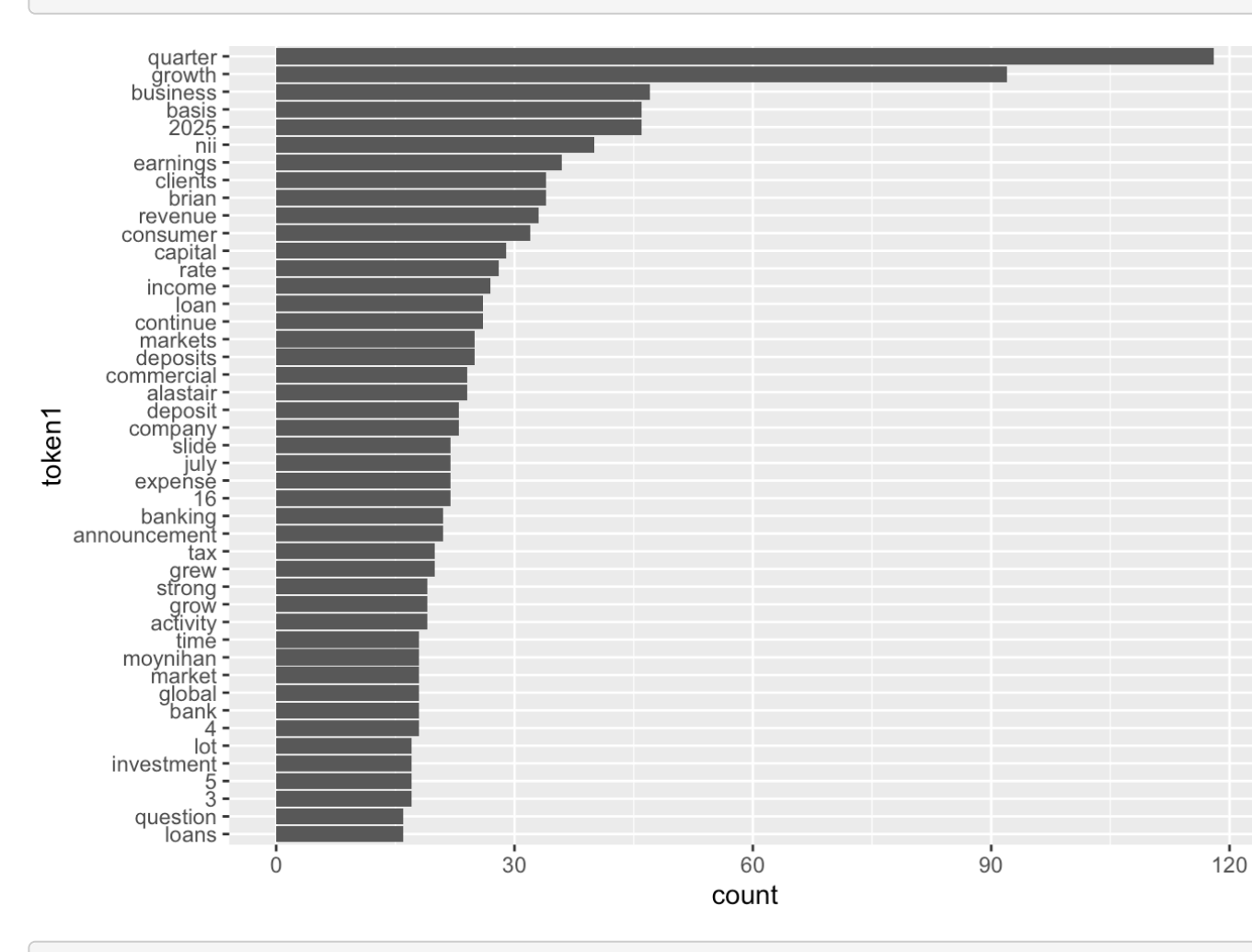
```r
view(stop.words)
```

```r
#remove stop words
token1 <- token1 %>%
  anti_join(stop.words, by =c('word'='X1'))
```

```r
token1 %>% nrow()
```

```
## [1] 4777
```

```r
token1 %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 15) %>%
  mutate(token1 = reorder(word, count)) %>%
  ggplot(aes(x=count , y=token1)) +
  geom_col()
```



```r
#look at similar words
arrange(token1,word) [1:500,1]
```

```
## # A tibble: 500 × 1
##    word
##    <chr>
##  1 0.89
##  2 1
##  3 1
##  4 1
##  5 1
##  6 1
##  7 1
##  8 1
##  9 1
## 10 1
## # ℹ 490 more rows
```

```r
#stem the tokens
stemmed <-token1 %>% mutate(stem=SnowballC::wordStem(word))
```

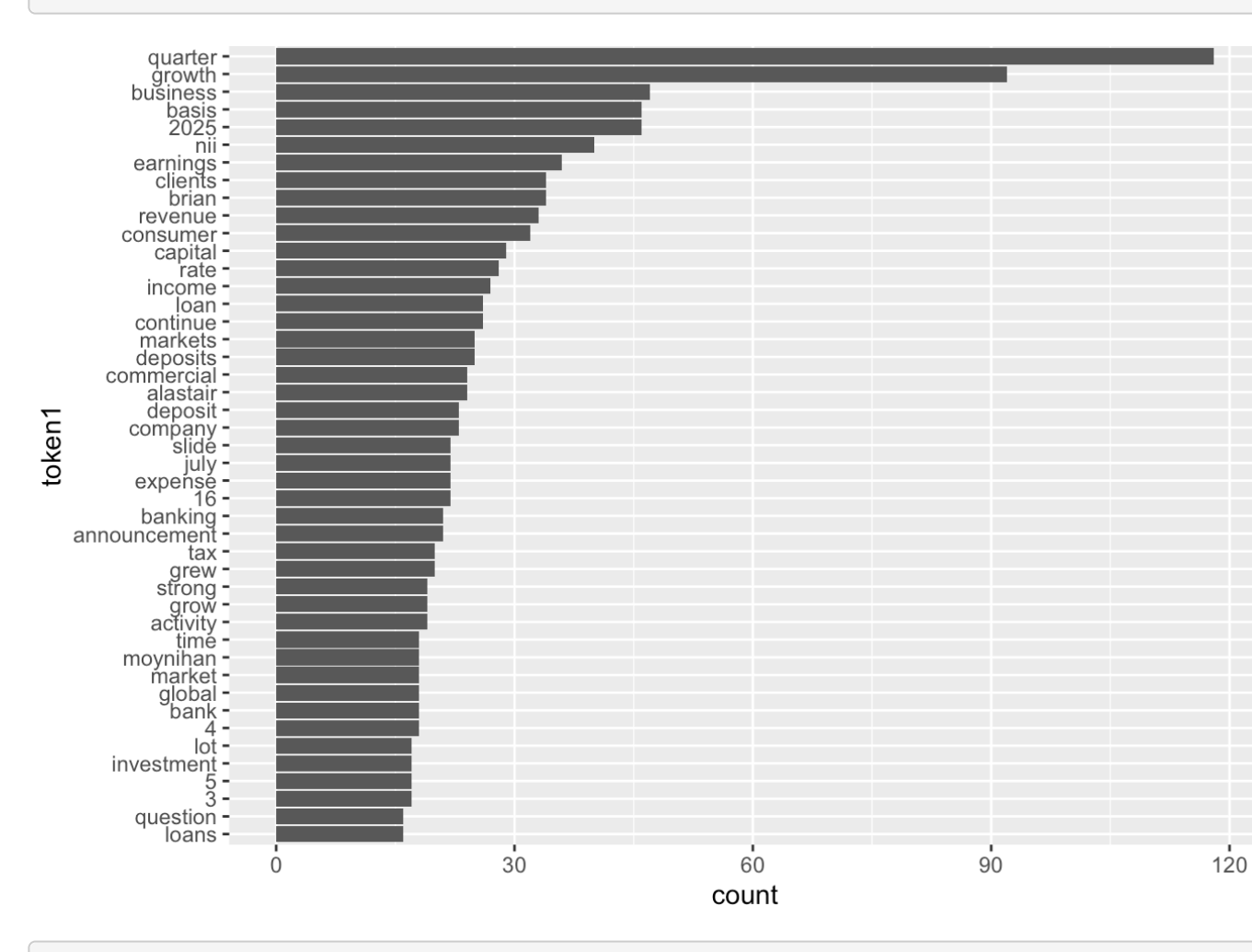
```r
view(stemmed)
```

```r
#look at similar words

arrange(stemmed,word)[1:500,1]
```

```
## # A tibble: 500 × 1
##    word
##    <chr>
##  1 0.89
##  2 1
##  3 1
##  4 1
##  5 1
##  6 1
##  7 1
##  8 1
##  9 1
## 10 1
## # ℹ 490 more rows
```

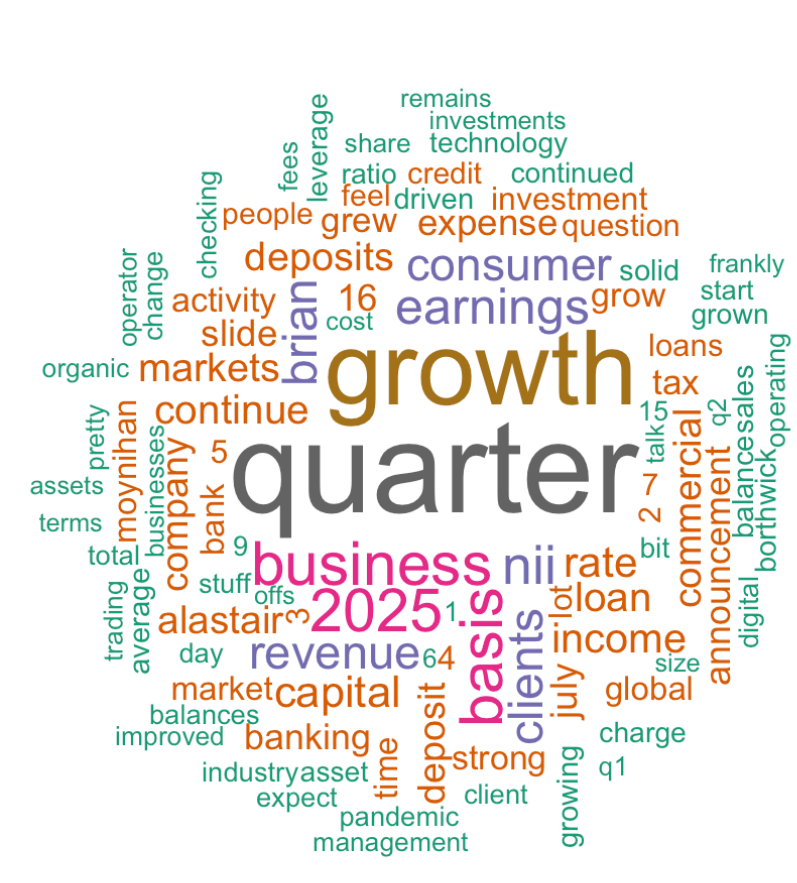
```r
stemmed %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  arrange(desc(count)) %>%
  filter(count > 15) %>%
  mutate(token1 = reorder(word, count)) %>%
  ggplot(aes(x=count , y=token1)) +
  geom_col()
```



```r
set.seed(77)
```

```r
library(wordcloud)
```

```r
stemmed %>%
  group_by(word) %>%
  summarize(count = n()) %>%
  with(wordcloud(words=word, freq=count, min.freq=10, max.words=120,
                 random.order = F, rot.per=0.30,
                 colors=brewer.pal(8,"Dark2")))
```



```r
#load finance sentiment list and explore
lm_dict<- tidytext::get_sentiments('loughran')
```

```r
#view dictionary
lm_dict %>% group_by(sentiment) %>% summarize(count=n())
```

```
## # A tibble: 6 × 2
##   sentiment    count
##   <chr>        <int>
## 1 constraining   184
## 2 litigious      904
## 3 negative      2355
## 4 positive       354
## 5 superfluous     56
## 6 uncertainty    297
```

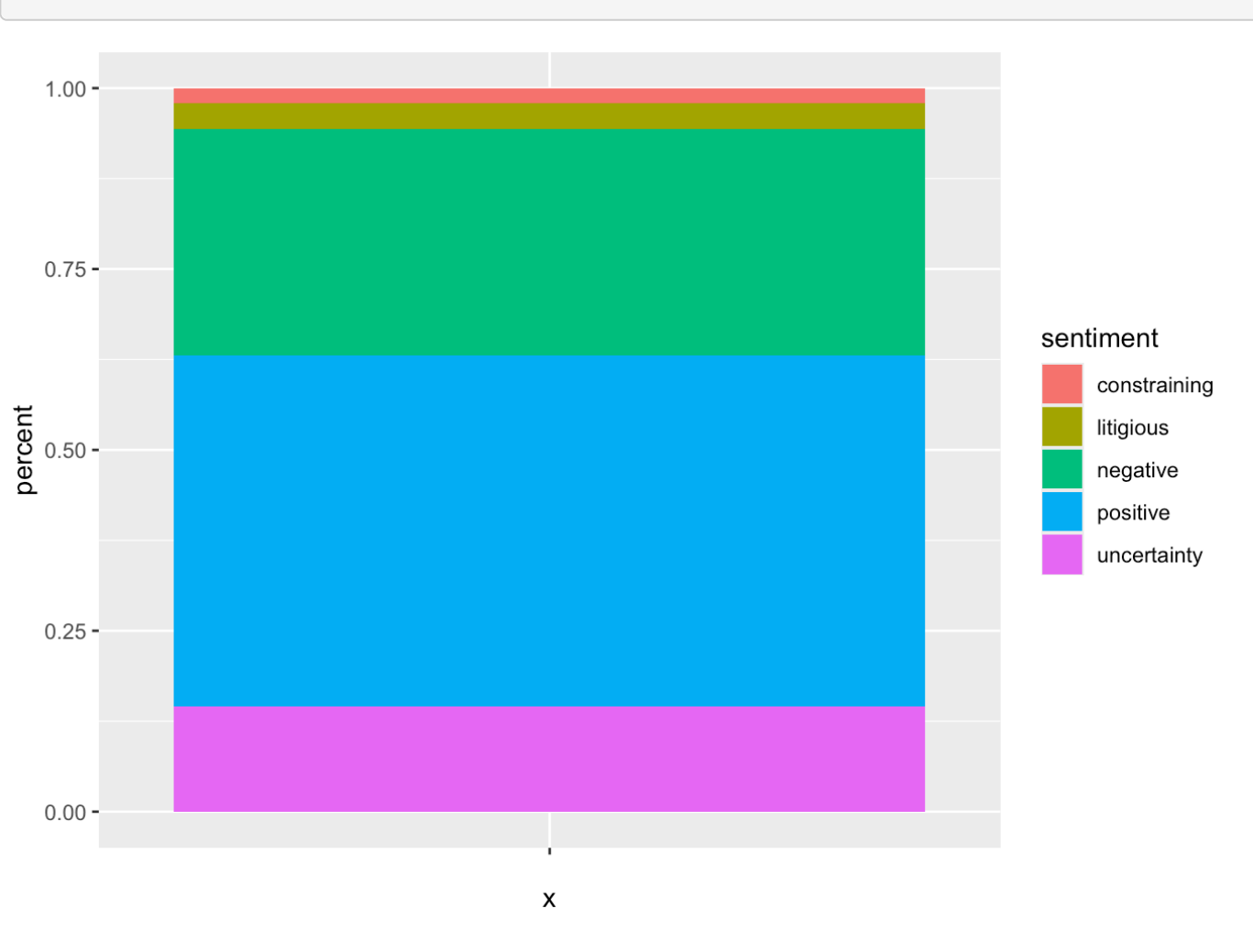
```r
#add sentiment
sentimented <- stemmed %>%
  inner_join(lm_dict, by='word')
```

```
## Warning in inner_join(., lm_dict, by = "word"): Detected an unexpected many-to-many relationship between `x` a
nd `y`.
## ℹ Row 1425 of `x` matches multiple rows in `y`.
## ℹ Row 2676 of `y` matches multiple rows in `x`.
## ℹ If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```r
#explore totals
sentimented %>%
  group_by(sentiment) %>%
  summarize(count =n (), percent = count/nrow(sentimented))
```

```
## # A tibble: 5 × 3
##   sentiment    count percent
##   <chr>        <int>   <dbl>
## 1 constraining     6  0.0213
## 2 litigious       10  0.0355
## 3 negative        88  0.312
## 4 positive       137  0.486
## 5 uncertainty     41  0.145
```

```r
sentimented %>%
  group_by(sentiment) %>%
  summarize ( count=n(), percent = count/nrow(sentimented)) %>%
  ggplot(aes(x='', y=percent, fill=sentiment)) +
  geom_bar(width=1, stat='identity')
```



Bank of America's NLP and Sentiment Analysis

Sentiment & Tone Breakdown *Positive (137 | 48.6%) Nearly half of the document carries a positive tone. This indicates strong emphasis on favorable outcomes, growth outlook, or confident language. Suggests management is focused on highlighting opportunities and successes.

*Negative (88 | 31.2%) About one-third of the content reflects negative sentiment. Likely tied to risks, challenges, or macroeconomic concerns. Balanced against the positive tone, it signals a realistic rather than overly optimistic communication.

*Uncertainty (41 | 14.5%) A moderate share of the language indicates uncertainty. Could point to future market conditions, regulatory changes, or forecasts that depend on external factors. This is fairly typical in financial communications where forward-looking statements require caution.

*Litigious (10 | 3.6%) References to legal or regulatory risks are limited. Suggests BAC is not heavily emphasizing ongoing disputes or litigation, but acknowledges compliance and oversight.

*Constraining (6 | 2.1%) Very minimal mentions of constraints. Indicates that financial or operational limitations are not a major theme in the communication.

The dominant positive tone (48.6%) outweighs the negative tone, projecting confidence and growth. Negativity and uncertainty together (~45%) show that management still addresses risks and unknowns, balancing optimism with realism.Litigious and constraining language are minor, suggesting legal risks and operational constraints are not pressing issues.

Takeaway: BAC's communication is overall optimistic with nearly half positive sentiment, but tempered with acknowledgment of risks and uncertainty. This mix can help build credibility with stakeholders, showing confidence without ignoring challenges.

Market Reaction Analysis to Bank of America's Q2 2025 Earnings Released, July 26,2025

Immediate Market Reaction (July 16, 2025 – Release Day) *BAC released results pre-market.* Stock closed slightly lower (–0.26%) at $46.03 despite beating earnings expectations. *The muted/negative reaction indicates that investors initially took a cautious stance, possibly due to mixed macro headlines or forward-looking risks mentioned in the report.

Investor Sentiment *Positive tone in management's outlook (capital allocation, strong balance sheet) likely supported the rally.* Market-wide backdrop was favorable, with the S&P 500 also trending up ~1.3% in the same period. *Compared to peers, BAC's stronger follow-through rally suggests investors viewed the results as a relative strength signal among large banks.

Takeaway *Day-of dip (Jul 16): cautious trading on release.* Follow-through rally (Jul 17–21): confidence returned, stock climbed steadily. *Net impact: Investors rewarded BAC's solid Q2 results, pushing shares up over 3% in the first 3 trading days after release, outperforming the broader market.

References:

Bank of America reports second quarter 2025 financial results. (n.d.). Bank of America. https://newsroom.bankofamerica.com/content/newsroom/press-releases/2025/07/bank-of-america-reports-second-quarter-2025-financial-results.html