1. "news_comment_BTS.csv"를 불러온 다음 행 번호를 나타낸 변수를 추가하고 분석에 적합하게 전처리하세요.

R 코드

```r
# Q1.
library(textclean)
news_comment <- raw_news_comment %>%
                mutate(id = row_number(), # 댓글자를 구분하기 위해 id 변수 생성
                reply = str_squish(replace_html(reply)))
news_comment
```

R 프로그램 결과

	reg_time	reply	press	title	url
1	2020-09-01 22:58:09	국보소년단	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
2	2020-09-01 09:56:46	아줌마가 들어도 좋더라	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
3	2020-09-01 09:08:06	팩트체크 현재 빌보드 HOT 100 1위 방탄소년단[BTS] 2위 Car...	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
4	2020-09-01 08:52:32	방탄소년단이 한국사람이라 너무 자랑스러워요 ㅠㅠ 우리오...	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
5	2020-09-01 08:36:33	대단한 BTS, 월드 클래스는 다르네^^ 좋은 소식!! 응원해요	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
6	2020-09-01 08:34:14	정국오빠 생일과 더불어 빌보드 1위기사라니ㅠㅠ축제구나	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
7	2020-09-01 08:32:14	정말 축하하고 응원하지만 집에서 여러 계정으로 스트리밍 ...	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
8	2020-09-01 08:22:09	기자는 자고 일어났지만, 팬들은 못자고 발표 기다림	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
9	2020-09-01 08:17:58	자랑스럽다!!!!!! 축하합니다!!!!	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
10	2020-09-01 08:15:37	SuperM 늘 응원하고 사랑합니다~	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
11	2020-09-01 08:15:14	빈집털이 아닌가 ㅠㅠ	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
12	2020-09-01 08:14:38	그런데 여러분 빌보드 들어가보시긴 한건가요? 도다체 어딜 ...	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
13	2020-09-01 08:14:07	마이클잭슨 마돈나나 하던 빌보드	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
14	2020-09-01 08:11:18	빌보드의 가치가 옛날에 비해 매우 떨어져서 아쉽네. 옛날엔 ...	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r
15	2020-09-01 08:10:38	화나요 25명은 일본인인가.. 너네 나라로돌아가~~	한국경제	[속보]BTS '다이너마이트', 한국 가수 최초로 빌보드 싱글 1위	https://news.naver.com/r

Showing 1 to 15 of 1,200 entries, 6 total columns

2. 댓글을 띄어쓰기 기준으로 토큰화하고 감성 사전을 이용해 댓글의 감정 점수를 구하세요.

R 코드

```
# Q2.
#토큰화
word_comment <- news_comment %>%
                unnest_tokens(input = reply, output = word,
                              token = "words", drop = F) %>%
                filter(str_length(word) > 1)
word_comment

# 감성점수 부여
word_comment <- word_comment %>%
                left_join(senti_dic, by = "word") %>% # 감성 사전 결합
                mutate(polarity = ifelse(is.na(polarity), 0, polarity)) # 감성 사전에 없으면 중립
word_comment %>% select(word, polarity)

# 감정 분류
word_comment <- word_comment %>%
                mutate(sentiment = ifelse(polarity == 2, "positive",
                                          ifelse(polarity == -2, "negative",
                                                 "neutural"))) %>% print()

# 댓글별 감정점수 생성
score_comment <- word_comment %>%
                 group_by(id, reply) %>%
                 summarise(score = sum(polarity)) %>%
                 ungroup()
```

R 프로그램 결과

	id	reply	score
1	1	국보소년단	0
2	2	아음마가 들어도 좋더라	0
3	3	엑트체크 현재 빌보드 HOT 100 1위 방탄소년단(BTS) 2위 Car...	0
4	4	방탄소년단이 한국사람이라 나무 자랑스러워요 ㅠㅠ 우리오...	0
5	5	대단한 BTS, 월드 클래스는 다르네^^ 좋은 소식!! 응원해요	4
6	6	청와오빠 생일과 더불어 빌보드 1위기사라니ㅠㅠ축제구나	0
7	7	정말 축하하고 응원하지만 집에서 여러 계정으로 스트리밍 ...	0
8	8	기자는 자고 일어났지만, 팬들은 못자고 청표 기다림	0
9	9	자랑스럽다!!!!! 축하합니다!!!!	2
10	10	SuperM 놀 응원하고 사랑합니다~	0
11	11	빈집털이 아닌가 ㅠㅠ	0
12	12	그런데 여러분 빌보드 들어가보시긴 한건가요? 도대체 어딜 ...	0
13	13	마이클잭슨 마돈나나 하던 빌보드	0
14	14	빌보드의 가치가 옛날에 비해 매우 떨어져서 아쉽네 옛날엔 ...	1
15	15	화나요 25명은 얼분인간. 너네 나라로돌아가~~	0
16	16	케이 팝중에 복미에서 매니아들많고 대중적으로 인지도- 민...	0
17	17	진짜.. 너무너무 자랑스럽다.. ㅠㅠ	2
18	18	축하합니다	0
19	19	빌보드1위는 세계대회 우승하고 맞먹는다 빌보드1위 정말위...	0

howing 1 to 19 of 1,188 entries, 3 total columns

3. 감정 범주 별 댓글 빈도를 나타낸 막대 그래프를 만드세요.

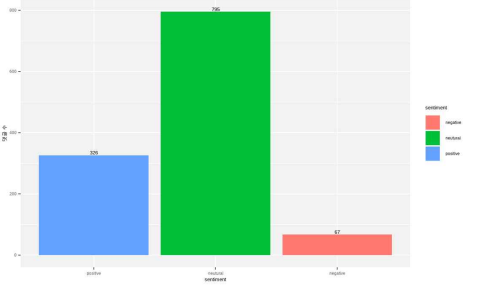
R 코드

```r
# Q3.
# 댓글의 감정분류
score_comment <- score_comment %>%
                 mutate(sentiment = ifelse(score >= 1, "positive",
                 ifelse(score <= -1, "negative", "neutural")))

# 댓글의 감정 빈도와 비율 생성
frequency_score <- score_comment %>%
                   count(sentiment) %>%
                   mutate(ratio = n/sum(n)*100) %>% print()

# 댓글의 감정 분류 막대 그래프 생성
frequency_score %>% ggplot(aes(x = sentiment, y = n, fill = sentiment)) +
                    geom_col() +
                    ylab("댓글 수") +
                    geom_text(aes(label = n), vjust = -0.3) +
                    scale_x_discrete(limits = c("positive", "neutural", "negative"))
```

R 프로그램 결과



4. 댓글을 띄어쓰기 기준으로 토큰화한 다음 감정 범주 별 단어 빈도를 구하세요.

R 코드

```
#Q4.
# 한글 단어 토큰화
comment <- score_comment %>%
           unnest_tokens(input = reply, output = word,
                         token = "words", # 단어 기준 토큰화
                         drop = F) %>%
           filter(str_detect(word, "[가-힣]") & # 한글만 추출
                  str_length(word) >= 2) %>% print()

# 감정 및 단어별 빈도 생성
frequency_word <- comment %>%
                  count(sentiment, word, sort = T) %>% print()

# 긍정 댓글 고빈도 단어
frequency_word %>% filter(sentiment == "positive")

# 부정 댓글 고빈도 단어
frequency_word %>% filter(sentiment == "negative")
```

R 프로그램 결과

sentiment	word	n
neutural	진짜	90
positive	진짜	88
neutural	빌보드	77
positive	자랑스럽다	77
positive	너무	71
neutural	방탄소년단	56
positive	정말	56
neutural	축하해요	52
neutural	군면제	50
neutural	정말	45

1-10 of 6,253 rows

sentiment	word	n
positive	진짜	88
positive	자랑스럽다	77
positive	너무	71
positive	정말	56
positive	방탄	40
positive	방탄소년단	39
positive	빌보드	36
positive	축하해	35
positive	대단하다	30
positive	좋은	22

1-10 of 2,113 rows

sentiment	word	n
negative	진짜	11
negative	너무	8
negative	빌보드	8
negative	힘든	7
negative	시기에	6
negative	국내	5
negative	군대	5
negative	군면제	5
negative	내가	5
negative	방탄소년단	5

1-10 of 870 rows

5. 로그 RR을 이용해 긍정 댓글과 부정 댓글에 상대적으로 자주 사용된 단어를 10개씩 추출하세요.

R 코드

```r
# Q5.
comment_wide <- frequency_word %>%
                filter(sentiment != " neutural") %>%
                pivot_wider(names_from = sentiment, # sentiment의 범주를 변수로 사용
                values_from = n, # 해당되는 값은 n으로 함
                            values_fill = list(n = 0)) %>% print()

# 로그상대위험 logRR
comment_wide <- comment_wide %>%
                mutate(log_RR = log(((positive + 1) / (sum(positive + 1))) /
                                    ((negative + 1) / (sum(negative + 1))))) %>% print()

# 로그 상대위험이 가장 큰 단어 10개씩 추출
top10 <- comment_wide %>%
         mutate(sentiment = ifelse(log_RR > 0, "positive", "negative")) %>%
         group_by(sentiment) %>%
         slice_max(abs(log_RR), n = 10, with_ties = F) %>% print()
```

R 프로그램 결과

word	neutural	positive	negative	log_RR	sentiment
<chr>	<int>	<int>	<int>	<dbl>	<chr>
국내	3	0	5	-2.119522	negative
없어서	0	0	5	-2.119522	negative
모르는	1	0	4	-1.937200	negative
있다	1	0	4	-1.937200	negative
날리	0	0	3	-1.714057	negative
폭도	2	0	3	-1.714057	negative
보다	2	0	3	-1.714057	negative
아니다	0	0	3	-1.714057	negative
없다	0	0	3	-1.714057	negative
케이팝	0	0	3	-1.714057	negative

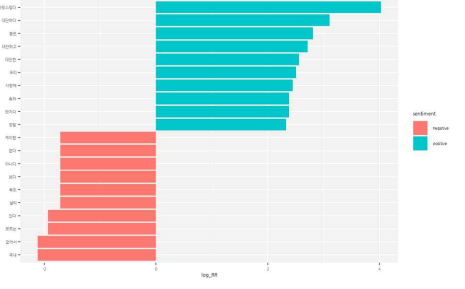
word	neutural	positive	negative	log_RR	sentiment
<chr>	<int>	<int>	<int>	<dbl>	<chr>
자랑스럽다	0	77	0	4.028946	positive
대단하다	1	30	0	3.106225	positive
좋은	1	22	0	2.807732	positive
대단하고	0	20	0	2.716760	positive
대단한	1	17	0	2.562609	positive
우리	14	16	0	2.505451	positive
사랑해	0	15	0	2.444826	positive
멋지다	0	14	0	2.380288	positive
축하	0	14	0	2.380288	positive
정말	45	56	3	2.328995	positive

6. 긍정 댓글과 부정 댓글에 상대적으로 자주 사용된 단어 각각 10개씩을 선택하여 긍정과 부정이 대비되도록 막대 그래프를 만드세요.

R 코드

```r
# Q6.
# 막대그래프 생성
top10 %>% ggplot(aes(x = reorder(word, log_RR), y = log_RR, fill = sentiment)) +
          geom_col() +
          coord_flip() +
          labs(x = NULL) +
          theme(text = element_text(family = "nanumgothic"))
```

R 프로그램 결과



7. 'Q3'에서 만든 데이터를 이용해 '긍정 댓글에 가장 자주 사용된 단어'를 언급한 댓글을 감정 점수가 높은 순으로 10개를 출력하세요.

R 코드

```
# Q7.
A <- score_comment %>% select(score, reply) %>%
                       arrange(-score) %>%
                       head(n=10) %>% print()
```

R 프로그램 결과

	score	reply
1	8	멋지다, 자랑스럽다, 대단하다 방탄소년단!!! 다이너마이트 ...
2	7	퍈은 아니야. 그래서 저 노력과 업적이 더 대단해보여. 정말 ...
3	7	ㅜㅜ. 진짜 이 코로나에 너희들이 빛이여. 핫백 1위라니. 모...
4	7	정국이 생일에 빌보드 핫100 1위라니... 정말 뜻깊은 하루네...
5	6	축하 합니다 우리에 보물이네 대한미국에 애국자 들이다 나...
6	6	축하 축하 아미분들도 축하^^
7	6	정말 대단하고 자랑스럽습니다.. 국적이 업그레이드 된거 같...
8	6	빌보드 핫100 1위 축하해요 여기까지 오느라 힘들었을텐데 ...
9	6	방탄소년단, 진짜 대단하고 대단하고 또 대단하다!! 무슨 말...
10	6	진짜 대단하다. K팝 아시아 최고 넘어서 빌보드 1위 등극 이...