

1. 트위터 크롤링, 텍스트 마이닝, 시각화를 위한 패키지 설치

**install.packages('twitteR')**

**install.packages('httr')**

**install.packages("base64enc")**

**install.packages("curl")**

**install.packages("rJava")**

**install.packages("KoNLP")**

**install.packages("wordcloud")**

**library(twitteR)**

**library(httr)**

**library(base64enc)**

**library(curl)**

**library(rJava)**

**library(KoNLP)**

**library(wordcloud)**

2. 트위터 API사용을 위한 OAuth(공개인증)을 위해 트위터 개발자앱 키와 토큰을 설정

**consumer\_key <- "(API key)"**

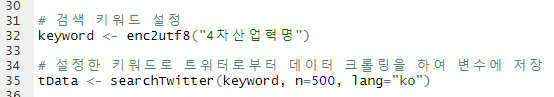
**consumer\_secret <- "(API secret key)"**

**access\_token <- "(Access token)"**

**access\_secret <- "(Access token secret)"**

3. 트위터 OAuth 공개인증 설정

**setup\_twitter\_oauth(consumer\_key, consumer\_secret, access\_token, access\_secret)**



4. 크롤링할 트위터 키워드를 설정

**keyword <- enc2utf8("검색 키워드 입력")**

5. 트위터로부터 데이터 크롤링을 실행하여 크롤링한 데이터들을 변수에 저장

**tData <- searchTwitter(keyword, n=500, lang="ko")** #500개의 한국어 트윗을 가져옴



6. 빅데이터를 담은 변수를 데이터 프레임으로 변환

**tData.df <- twListToDF(tData)**

7. 크롤링 한 데이터를 텍스트 파일로 저장

**write(tData.df$text, "tData.txt")**

8. 텍스트 파일을 한 줄 씩 읽어들여 변수에 저장

**twitterPosts <- readLines("tData.txt")**

9. 해당 데이터에서 명사만을 추출

**twitterWords <- sapply(twitterPosts,extractNoun,USE.NAMES=F)**

10. 명사만을 추출한 데이터에서 추가로 점, 숫자, 빈공간, 그 외의 불필요한 단어들을 제거

**interest<-gsub("[[:punct:][:digit:][:space:]]", "", unlist(twitterWords))**

**interest<-gsub("https", "", unlist(interest))**

**interest<-gsub("@", "", unlist(interest))**

11. 정제된 텍스트 데이터를 텍스트 파일로 저장

**write(interest, "twinterest.txt")**

12. 단어 길이가 3개 보다 큰 단어만을 분류

**interest <- Filter(function(x) {nchar(x) > 3}, interest)**

13. 단어의 빈도 분석

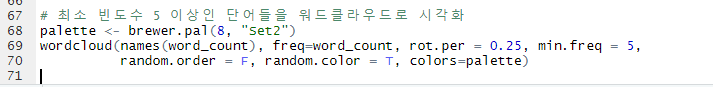
**word\_count <- table(interest)**

14. 빈도수 상위 30개를 출력

**head(sort(word\_count, decreasing = T), 30)**

15. 단어의 빈도수를 담은 데이터를 텍스트 파일로 저장

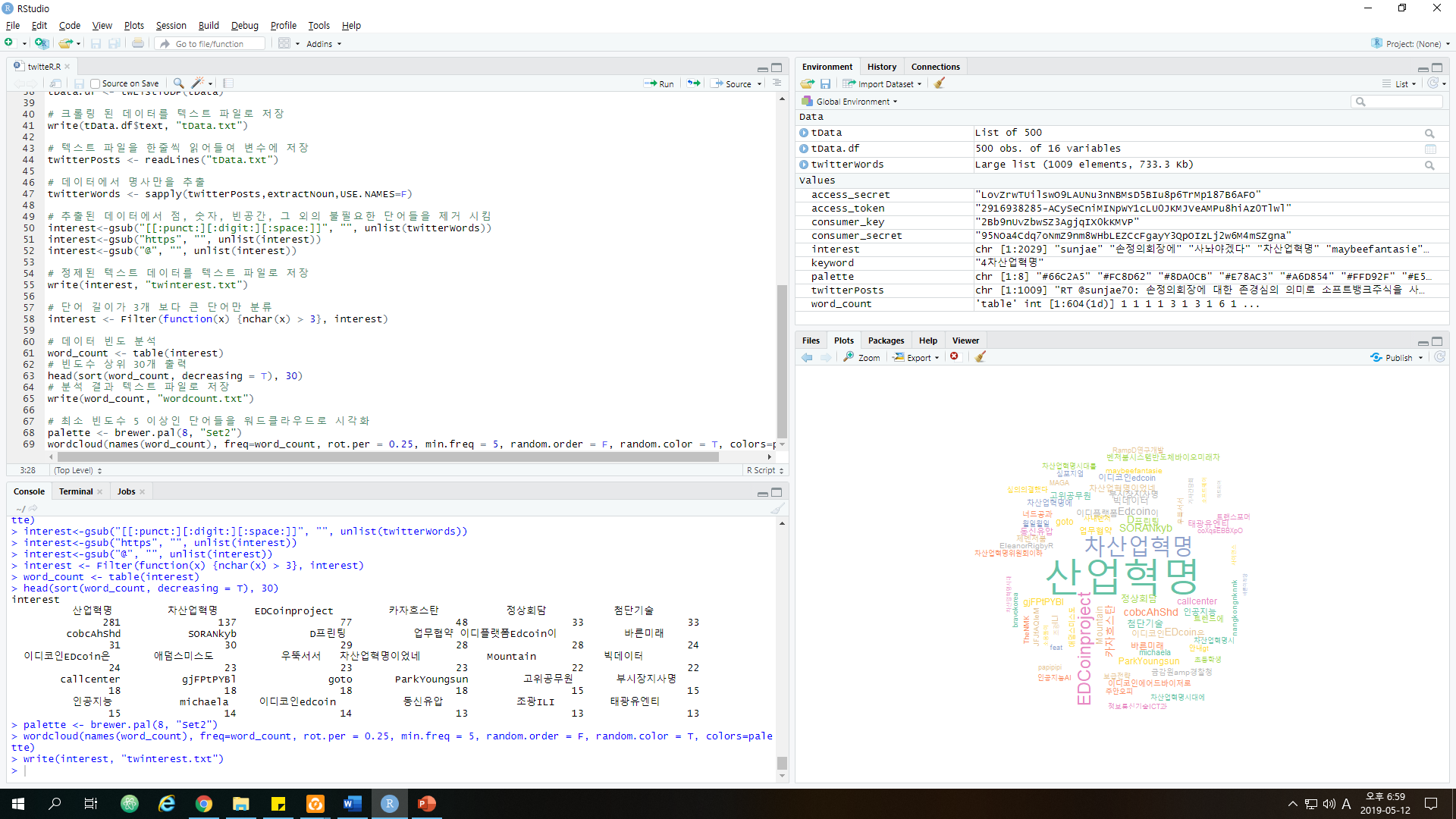
**write(word\_count, "wordcount.txt")**



16. 최소 빈도수 5 이상인 단어들을 워드 클라우드로 시각화

**palette <- brewer.pal(8, "Set2")**

**wordcloud(names(word\_count), freq=word\_count, rot.per = 0.25, min.freq = 5, random.order = F, random.color = T, colors=palette)**



완성된 결과물