트위터 분석

해쉬태그 키워드 중심

2020-01-26

본 보고서는 [다양한 데이터 가져오기 - “트위터 데이터 과학”](https://statkclee.github.io/ingest-data/ingest-twitter-data-science.html) 웹문서를 Makefile을 작성하여 자동화하는 것을 재현하는 과정을 담고 있다.

# 트위터 데이터

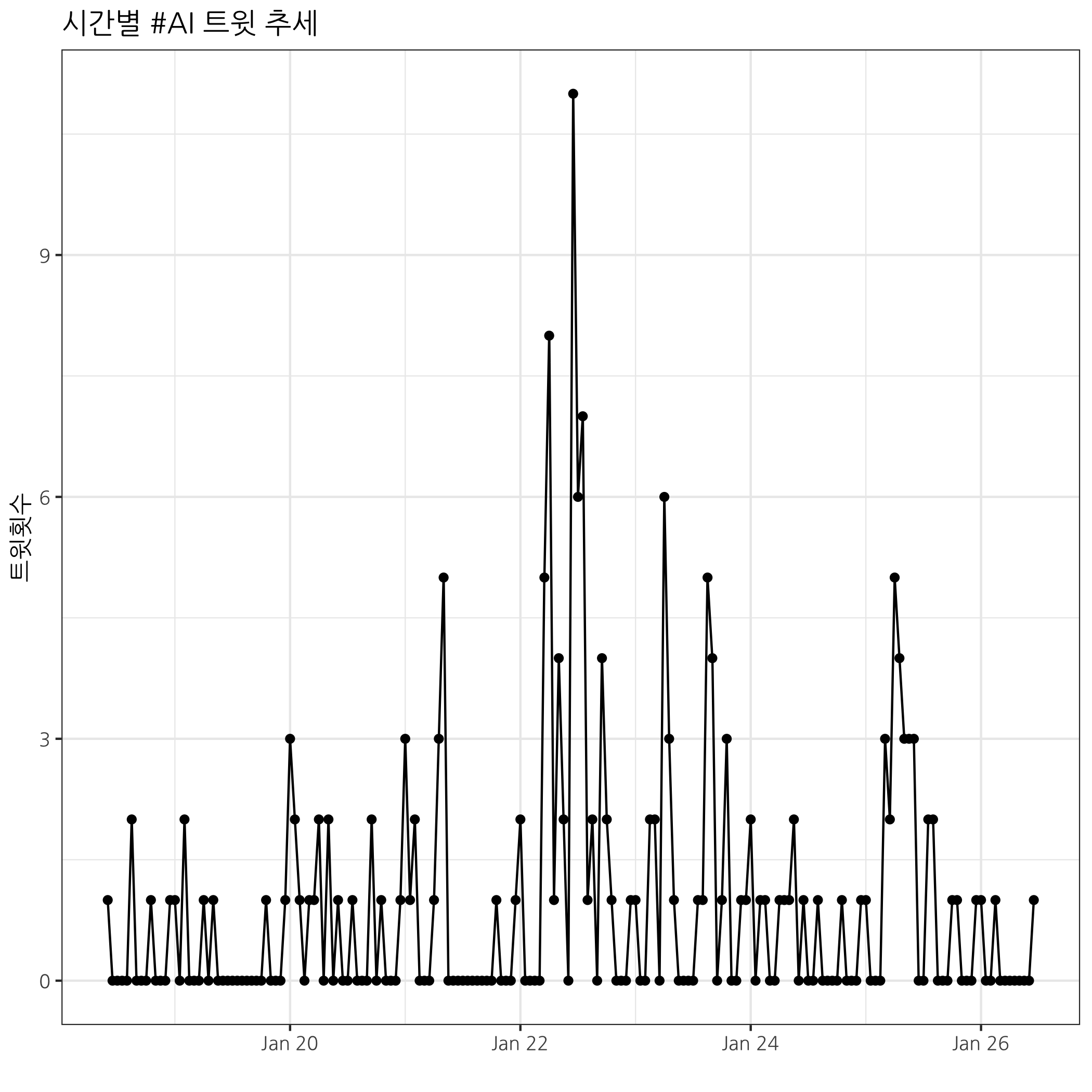
가장 먼저 트위터 데이터를 가져와야 한다. 이를 위해서 get\_data.R 스크립트에 내장된 get\_data() 함수를 사용해서 트위터에서 데이터를 가져와서 이를 내부적으로 저장시킨다. 원래 DT 팩키지를 사용해서 인터랙티브하게 분석 보고서를 작성하는 것이 좋으나, .pdf, .docx 파일 출력결과를 생각해서 kable() 함수를 사용해서 이를 정적 문서에서도 가시성이 좋도록 표를 작업한다. kable() 함수가 PDF, Word 파일 생성시 표를 제대로 만들어 내지 않는 경우 pander()로 대신한다.

library(tidyverse)  
library(pander)  
  
panderOptions('table.alignment.default', function(df)  
 ifelse(sapply(df, is.numeric), 'right', 'left'))  
panderOptions('table.split.table', Inf)  
panderOptions('big.mark', ",")  
panderOptions('keep.trailing.zeros', TRUE)  
  
twit\_df <- read\_rds("data/tw\_dat.rds")  
  
twit\_df %>%   
 sample\_n(10) %>%   
 select(screen\_name, created\_at, text) %>%   
 pander()

|  |  |  |
| --- | --- | --- |
| screen\_name | created\_at | text |
| busk2019 | 2020-01-22 11:38:05 | ‘#공정’ ‘#경제’ 이어 ‘#AI’ #안철수, 카이스트 방문해 AI국가전략 경청 AI 비전 논의 기술·발전 상응하는 법·제도 방향  <https://t.co/udYNxsYkPG> |
| AI\_bot\_ | 2020-01-21 02:00:58 | 「 피식거려도 이해해줄수있겠지-…? 」#AI |
| hahah\_2020 | 2020-01-23 06:08:27 | #안철수 #안전’에 굉장이 관심이 많다. 사람이 일일이 모니터링하는 방식은 한계가 있다”면서 “#AI 기술을 이용해 사람들의 안전을 획기적으로 높일 수 있다 정부가 지금부터 적극적으로 열심히 투자해서 AI인재 10만명 양성 방안을 가급적 빠르게 마련해주길 바란다 <https://t.co/1vNhdZZI70> |
| 4evermaya1 | 2020-01-22 06:15:13 | 국민 사이다 되시는 거 아님?ㅎ 답답하게 쌓인 국민들 마음 시원하게 뚫릴 듯.. #공감맨 #개념남 #백신맨 #안철수 #안철수박사 #AI #경제전문가 #교수 #의사 #러너 #안보전문가 #교육전문가 #천재 안철수, 문재인에 일침…“한 번도 경험해보지 못한 무능” <https://t.co/Ba0aKyExWi> |
| Frei\_AberEinsam | 2020-01-23 15:06:53 | #안철수\_미래세대를\_위한\_프로젝트 나는 우리 미래세대를 위한 러닝5.0 프로젝트를 이미 개발하기 시작했다 #다시안철수🏃🏻‍♂️ #미래세대 #인공지능 #AI #러닝5.0 <https://t.co/WTndzJGyIg> |
| AI\_bot\_ | 2020-01-20 08:30:55 | 「 채색이 망가진 세계 안에서 아직 어린 눈에 찌른 한 송이의 붉은 장미」#AI |
| Eueukim1 | 2020-01-25 05:09:36 | 환희 말 못알아듣는 A.I #업텐션 #up10tion #미치게해 #AI <https://t.co/OMjeujyGFP> |
| AI\_bot\_ | 2020-01-21 08:30:54 | 「 채색이 망가진 세계 안에서 아직 어린 눈에 찌른 한 송이의 붉은 장미」#AI |
| l8ueYa8GPKaUfqh | 2020-01-22 13:02:49 | ‘공정’ ‘경제’ 이어 ‘#AI’. 안철수, 카이스트 방문해 AI국가전략 경청. <https://t.co/mhmJBi6xO5> ●23일 4차 산업혁명시대의  AI 비전 논의. ●기술·발전 상응하는 법·제도 방향 생각 밝힌다. ●“AI가 우리 삶 전체와 국가경쟁력을 좌우하게 될 것” ●미래 먹거리·미래 일자리 위한 ‘일하는 정치’ 언급 <https://t.co/0wXXFVNLjx> |
| Frei\_AberEinsam | 2020-01-23 15:09:49 | “누가 수구이고 누가 개혁인가는 간단하다. 오늘 할 수 있는 개혁을 내일로 미루는 사람이 수구다.” #안철수(20170109)🌸 #안파고 #AI <https://t.co/6p4M5Mu3wI> |

# 트위터 추세

graph\_trend.R 파일에 graph\_trend() 함수를 사용해서 twit\_trend.png 파일을 생성하여 특정 해쉬태그 트위터 데이터의 추세를 파악하고자 시각화하느 결과를 보여주는 것이 목적이다.

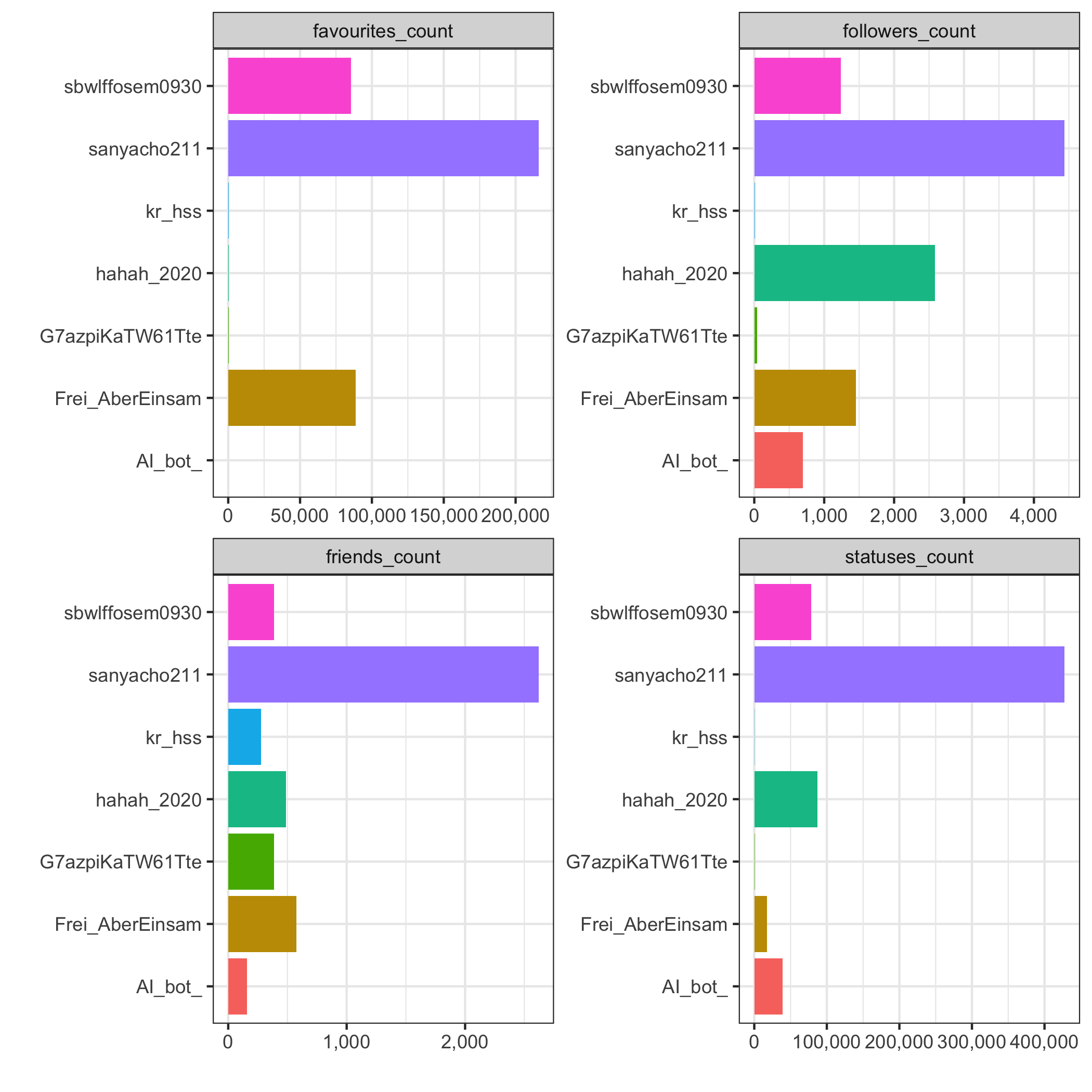


트위터 해쉬태그 추세

# 사용자 활동성 지표

특정 해쉬태그를 많이 언급한 사용자에 대한 기본 프로파일 정보를 lookup\_users() 함수를 통해서 파악할 수 있고 이를 바탕으로 다음 사용자 프로파일 활동성 정보를 얻어올 수 있다.

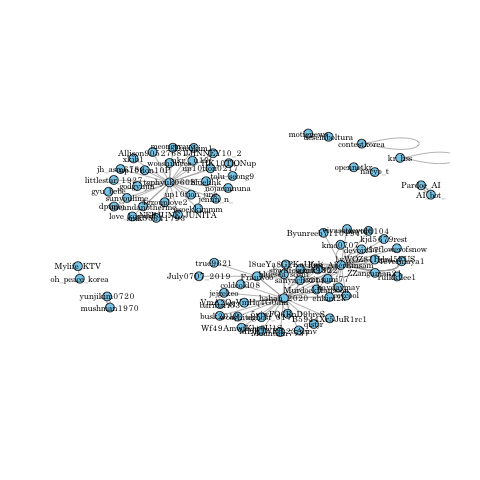
* status
* followers
* friends
* favourites



트위터 해쉬태그 추세

# 리트윗 네트워크 분석

특정 해쉬태그 관련 트윗 검색어를 넣어 include\_rts = TRUE 리트윗을 추가하게 되면 리트윗(retweet) from, to 자료구조를 통해서 네트워크 분석을 통해서 트윗이 어떻게 전파되는지, 전파경로를 파악할 수 있다.

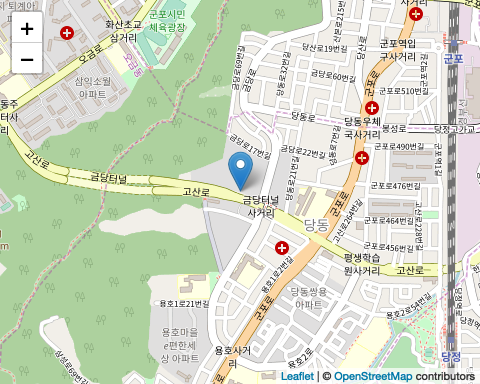


리트윗 네트워크 분석

# 공간정보

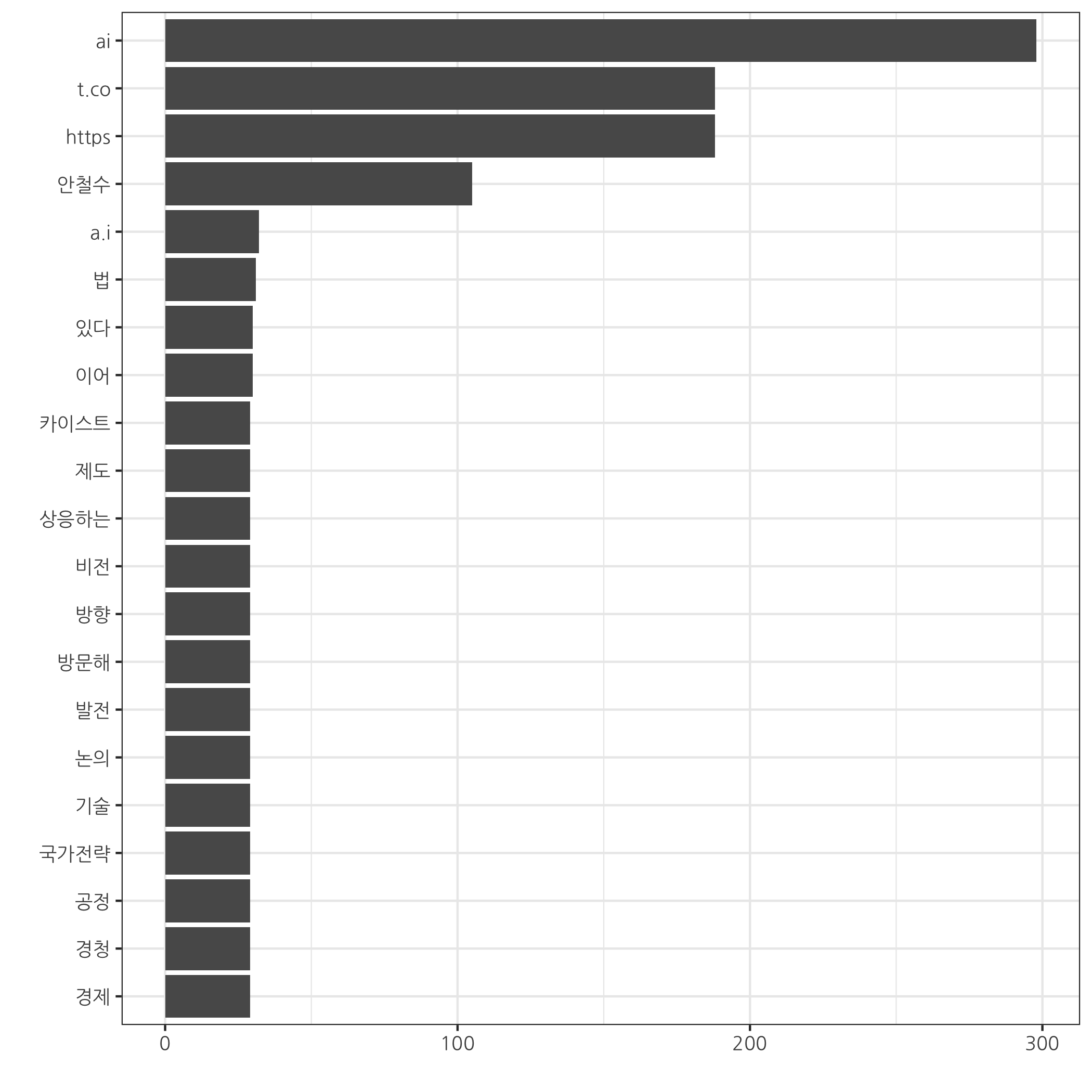
특정 해쉬태그 관련 트윗 검색어를 넣어 어디에서 많이 하는지 공간정보가 포함되어 있는 경우 이를 반영시킨다.

library(rtweet)  
library(leaflet)  
  
geo\_twts\_df <- lat\_lng(twit\_df) %>%   
 select(screen\_name, text, lng, lat) %>%   
 filter(!is.na(lng))  
  
  
geo\_twts\_df %>%   
 leaflet() %>%   
 addProviderTiles(providers$OpenStreetMap) %>%   
 addMarkers(lng=~lng, lat=~lat, clusterOptions = markerClusterOptions(),  
 popup = ~ as.character(paste0("<strong>", paste0("유저명: ",`screen\_name`), "</strong><br>",  
 "-----------------------------------------------------------<br>",  
 "&middot; 트윗: ", text)))



# 트윗 텍스트 분석

특정 해쉬태그관련 트윗 검색어를 텍스트 데이터로 놓고 이를 자연어처리 분석을 추진한다. 이를 위해서 tidytext 팩키지를 통해 단어 토큰을 만들이고 불용어 처리한 후에 단어 빈도수를 시각화해 보자.



트윗 텍스트 분석