

# 트위터 분석

## 해쉬태그 키워드 중심

true

2020-01-26

## Contents

트위터 데이터	1
트위터 추세	2
사용자 활동성 지표	2
리트윗 네트워크 분석	5
공간정보	5
트윗 텍스트 분석	6

본 보고서는 다양한 데이터 가져오기 - “트위터 데이터 과학” 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 패키지를 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 pandoc()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
daehannews_kr	2020-01-21 08:36:42	가짜영상과의 전쟁 #가짜영상 #가짜뉴스 #음란물 #딥페이크 #포르노 #AI #지인능육 #처벌 #이미지기반성적학대 #카드뉴스 <a href="https://t.co/H5NmSoklkV">https://t.co/H5NmSoklkV</a>
yunjikim0720	2020-01-20 00:33:50	#GS25'가 #BC카드 #스마트로'와 함께 #AI 기반 무인 편의점을 선보였습니다. 작년에는 #이마트24'도 뛰어드는 등 국내도 무인점포 본격화 #amazonGo <a href="https://t.co/ITi7scNBjH">https://t.co/ITi7scNBjH</a> <a href="https://t.co/UabKVeEXb5">https://t.co/UabKVeEXb5</a>
AI_bot_	2020-01-19 08:30:54	「채색이 망가진 세계 안에서 아직 어린 눈에 찌른 한 송이의 붉은 장미」#AI
AI_bot_	2020-01-23 13:30:55	「채색이 망가진 세계 안에서 아직 어린 눈에 찌른 한 송이의 붉은 장미」#AI
dptmej_	2020-01-25 14:16:33	환희 말 못알아듣는 A.I #업텐션 #up10tion #미치게해 #AI <a href="https://t.co/OMjeuJyGFP">https://t.co/OMjeuJyGFP</a>
AI_bot_	2020-01-24 00:00:56	「감정회로 Error 라니....?」#AI
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 <a href="https://t.co/WTndzJGylg">https://t.co/WTndzJGylg</a>
hahah_2020	2020-01-23 15:16:35	#안철수 미래세대를 위한 프로젝트 나는 우리 미래세대를 위한 러닝5.0 프로젝트를 이미 개발하기 시작했다 #다시안철수☑ #미래세대 #인공지능 #AI #러닝5.0 <a href="https://t.co/WTndzJGylg">https://t.co/WTndzJGylg</a>
blueseascent	2020-01-22 17:07:06	‘#공정’ ‘#경제’ 이어 ‘#AI’ #안철수, 카이스트 방문해 AI국가전략 경청 AI 비전 논의 기술·발전 상응하는 법·제도 방향 <a href="https://t.co/udYNxsYkPG">https://t.co/udYNxsYkPG</a>
AI_bot_	2020-01-21 19:15:55	「기대한다고? 그런말 안믿어」라고해놓고 바보같이 믿고있는 한소녀 #AI

## 트위터 추세

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

## 사용자 활동성 지표

특정 해시태그를 많이 언급한 사용자에게 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

시간별 #AI 트윗 추세

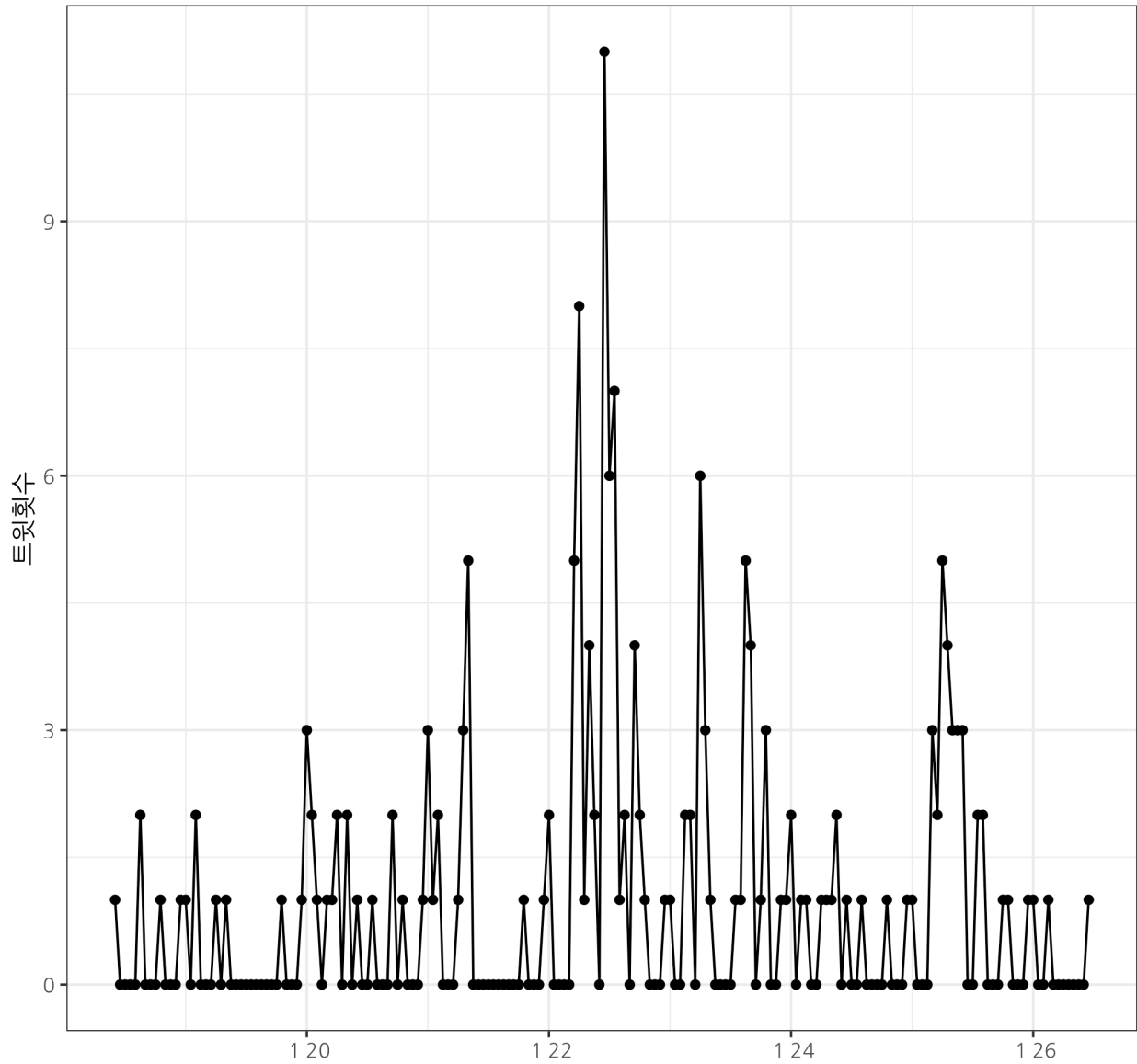


Figure 1: 트위터 해시태그 추세

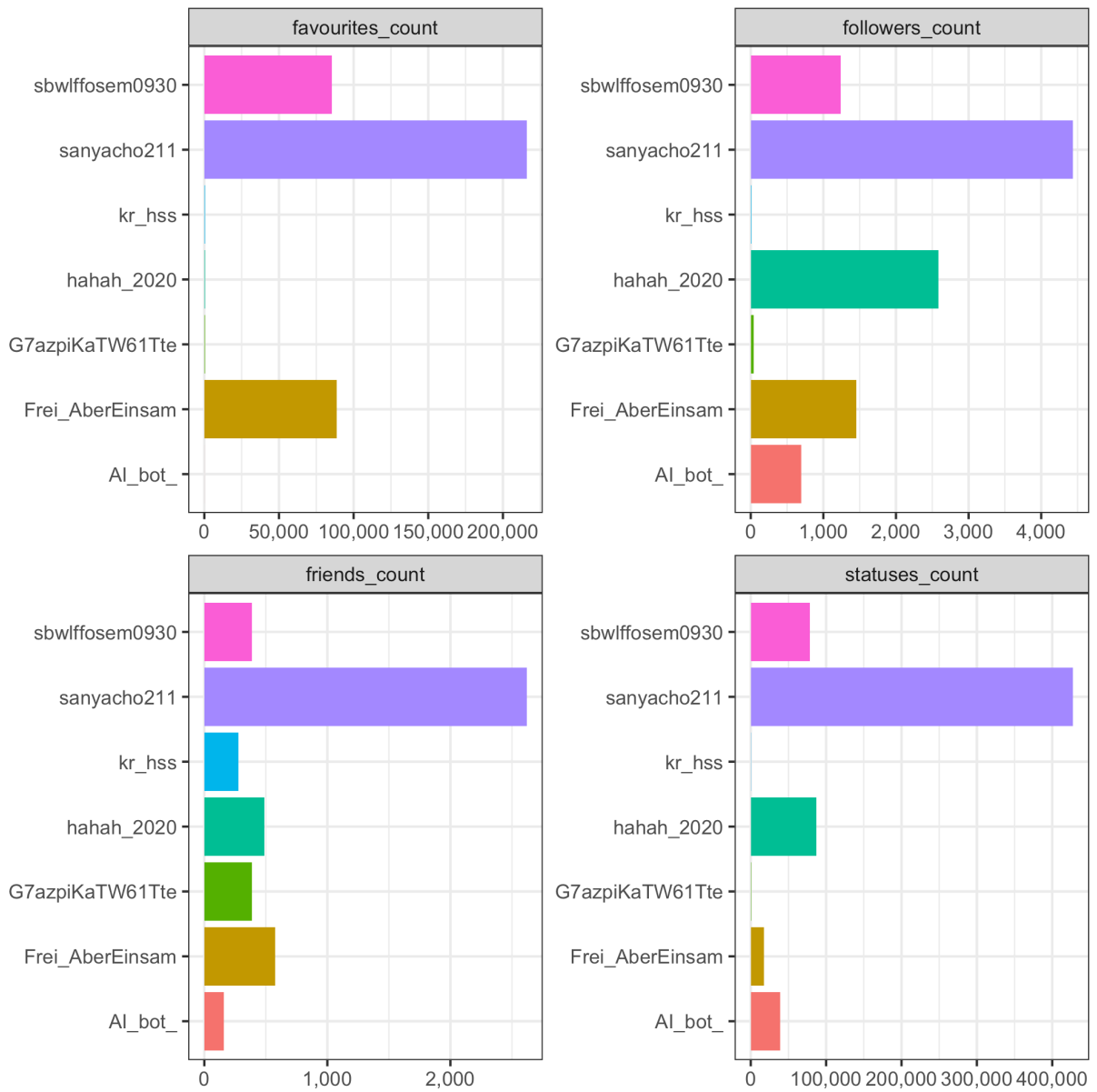


Figure 2: 트위터 해시태그 추세

## 리트윗 네트워크 분석

특정 해시태그 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.

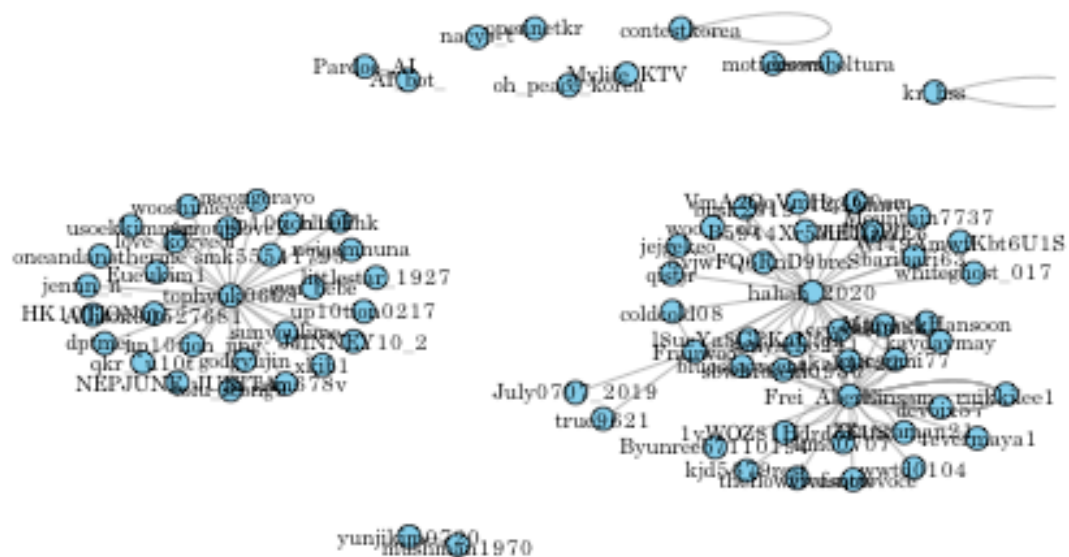


Figure 3: 리트윗 네트워크 분석

## 공간정보

특정 해시태그 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(" : ", `screen_name`), "</strong><br>",
      "-----  

      "&mdot; : ", text)))

```



## 트윗 텍스트 분석

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

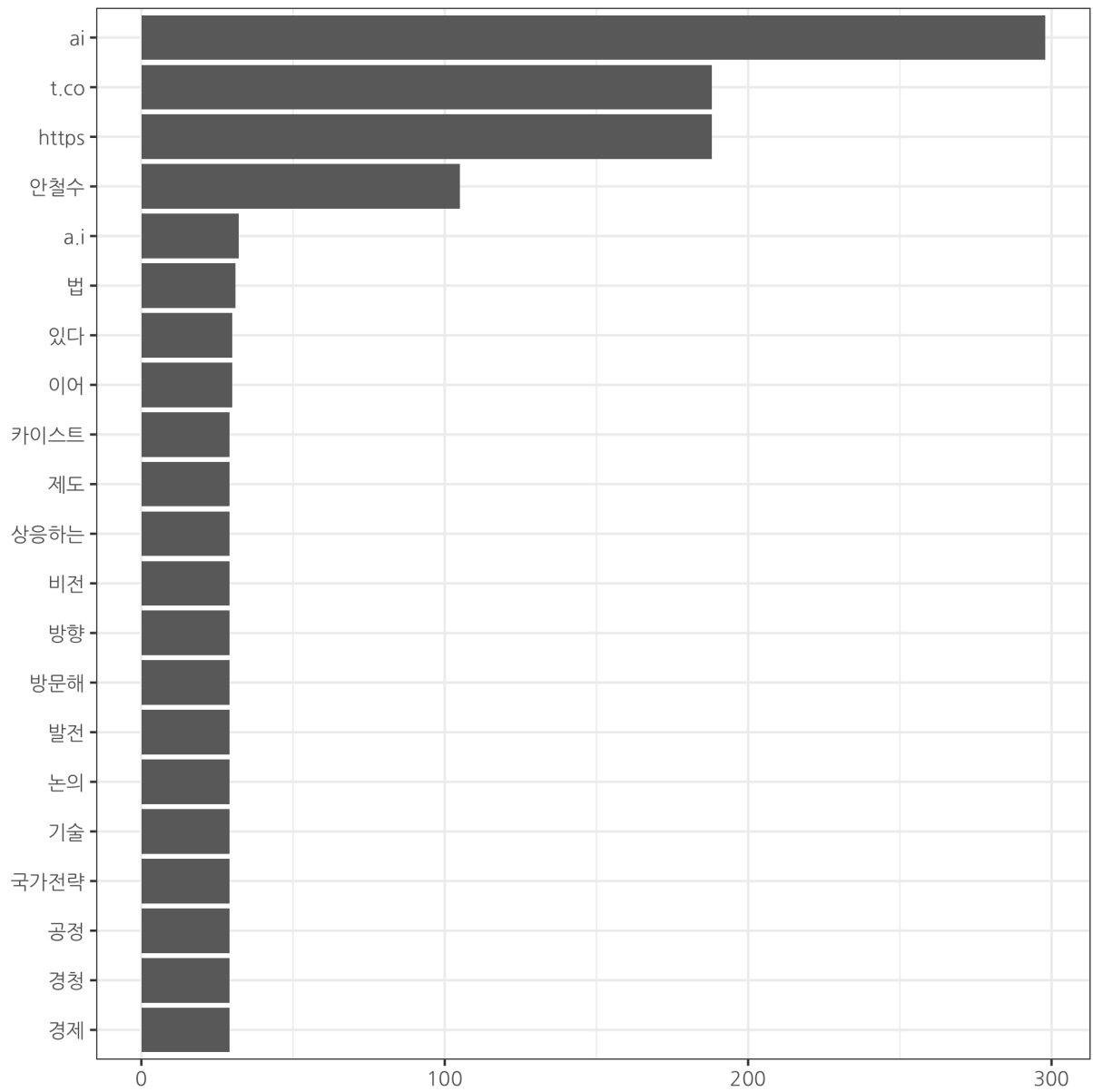


Figure 4: 트윗 텍스트 분석