# WORKSHOP: DIGITAL TRACE DATA

Taehee Kim
University of Oldenburg

Dec 9th-10th 2021

# TAEHEE KIM

- Post-doc researcher
- Ph.D. Political Science, Keio University, Japan
- M.Sc. Social and Economic Data Anlysis, University of Konstanz
- Hate speech spread in social media
- Conspiracy theory and extremists' movement in social media

**Digital trace data workshop**

- Gesis 2020 (R)
- ECPR winter school 2019, 2020 (Python)

# THE DIGITALIZATION AND DATA REVOLUTION FOR THE SOCIAL SCIENCES

- Nowadays, diverse kinds of digital trace data with a large volume have become available to researchers.
- It has a great potential for new approaches to social science questions.
- Vast textual corpora
- Alternatives to telephone or face-to-face surveys
- Users’s interactions with various digital devices or services

# AVAILABLE SOURCES

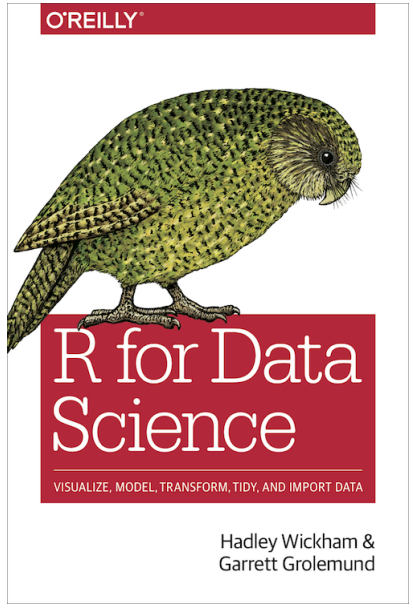
- Twitter: APIs are available, open with some restrictions. Provides rich data, text, network, images, geo tag etc.
- Facebook: Largely restricted unless you work with them.
- Instagram: Great for research using images. You need to scrape.
- Youtube: APIs are available, open.
- Online communities, e.g., reddit, Wikipedia etc..

This workshop introduces Twitter API and provides multiple examples of collecting and analyzing Twitter data in R. More concretely we work on extracting meaningful information from Twitter data, exploring retweet network, conducting simple text analysis.

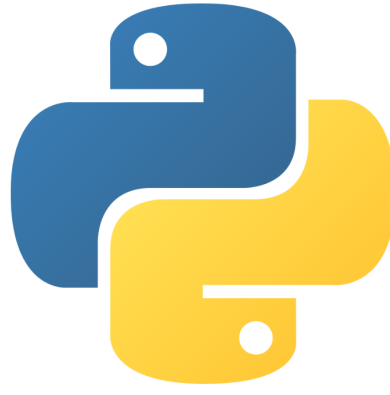
# THE LANGUAGE

...with **rtweet** package.

# PREPARATION




- R and RStudio in your laptop
- Base R + tidyverse
- Wickham and Grolemund. 2017. R for Data Science. O'Reilly, especially **Data transformation** chapter.
- For other self-learning resources, check Github rep.



In case you want to use Python in the future..

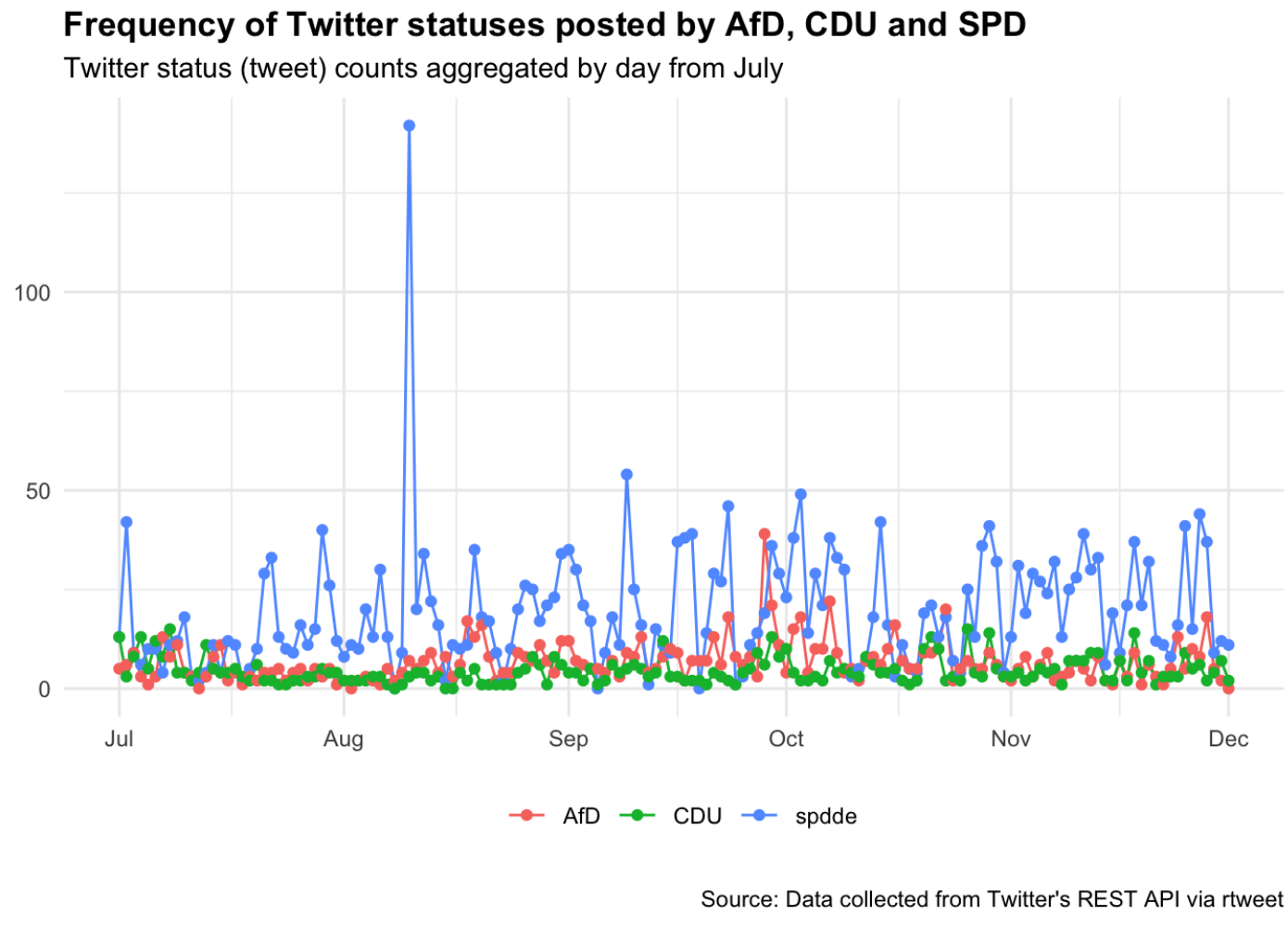
Juergens, Pascal and Andreas Jungherr. 2016. “A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis. SSRN2710146


- Python for data collection and management
- R for analysis and visualization
- Juergens, Pascal and Andreas Jungherr. 2016.
- ZC Steinert-Threlkeld. 2017. Twitter as Data. Cambridge University Press

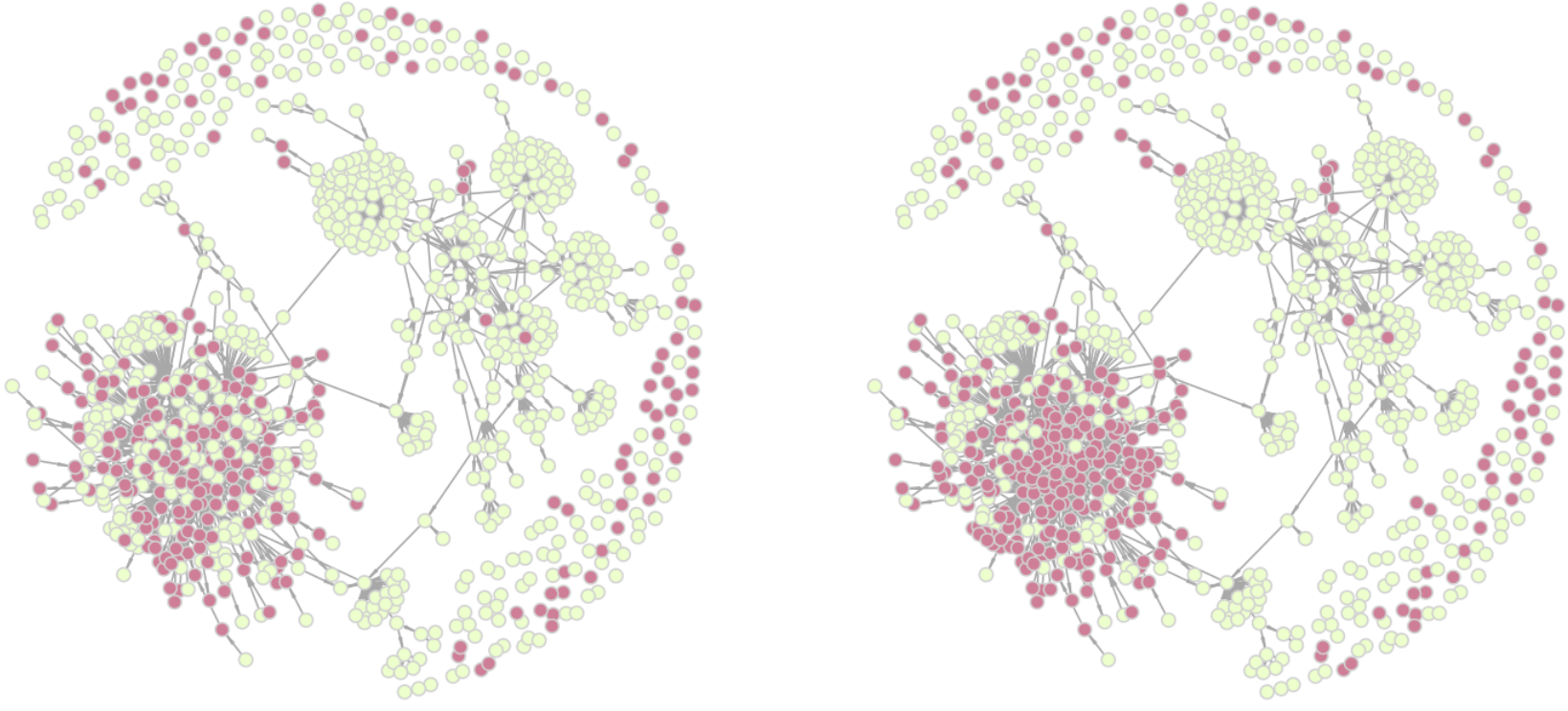
# TWO DAY (12 HOURS) WORKSHOP
# 9TH-10TH DECEMBER 2021

| | 17th | 18th |
|---|---|---|
| **9:00 - 10:30** | Session 1 | Session 5 |
| **10:45 - 12:15** | Session 2 | Session 6 |
| Lunch Break | | |
| **13:00 - 14:30** | Session 3 | Session 7 |
| **14:45 - 16:15** | Session 4 | Session 8 |

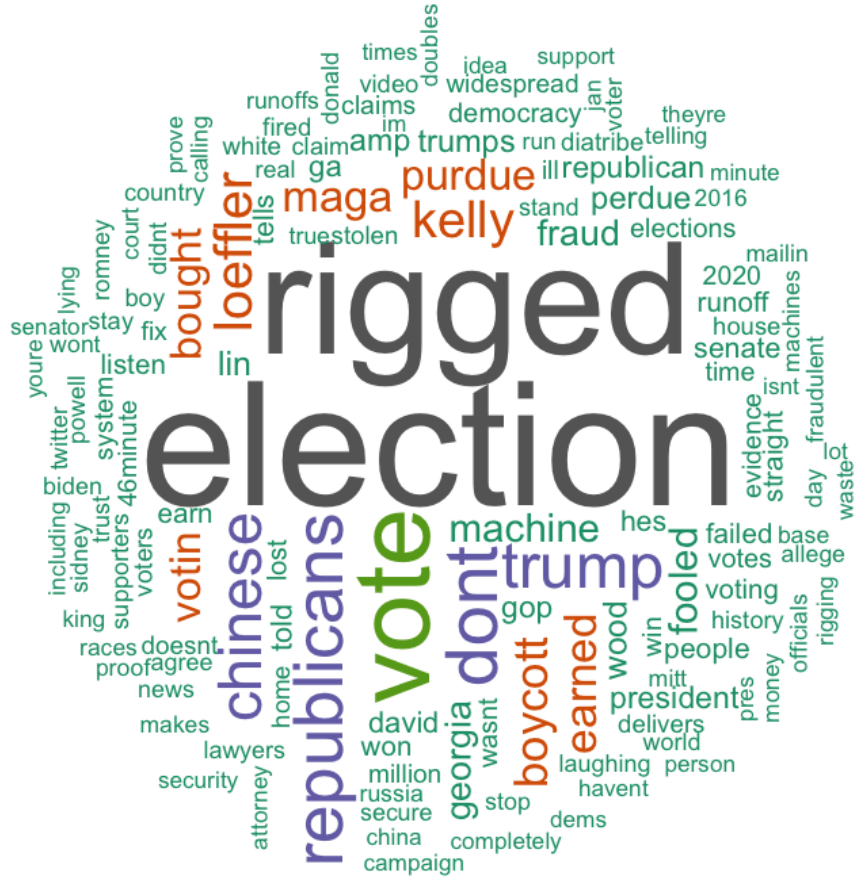
Session 1. Introduction to Twitter APIs
Session 2. Collecting Tweets 1: Rest APIs
Session 3. Collecting Tweets 2: Streaming APIs




# Session 4-5. Exploring Twitter data (summary statistics)



Session 6-7. Twitter user retweet network: Centurality, Cluster detection




# Session 8. Simple text analysis: Word cloud

# WORKSHOP GITHUB REPOSITORY

You can download materials from here:

https://github.com/thkim0321/gesis2021digitaltrace

# TWITTER'S API

*" APIs are the way computer programs talk to each other so that they can request and deliver information. ( About Twitter's APIs) "*

**You need to create Twitter applications, so called ‘apps’, which you need to access Twitter APIs. There are two options for you...**

1. Prepare your own developer account
2. Send me your @handle of Twitter account More detail