



Collins Dictionary named Climate Strike the word of 2019:

"a form of protest in which people absent themselves from education or work in order to join demonstrations demanding action to counter climate change"

"Climate Strike and public discourse" is the topic of my thesis.

In my research I've encountered two issues with gathering empirical data from public discourses around Climate Strike.

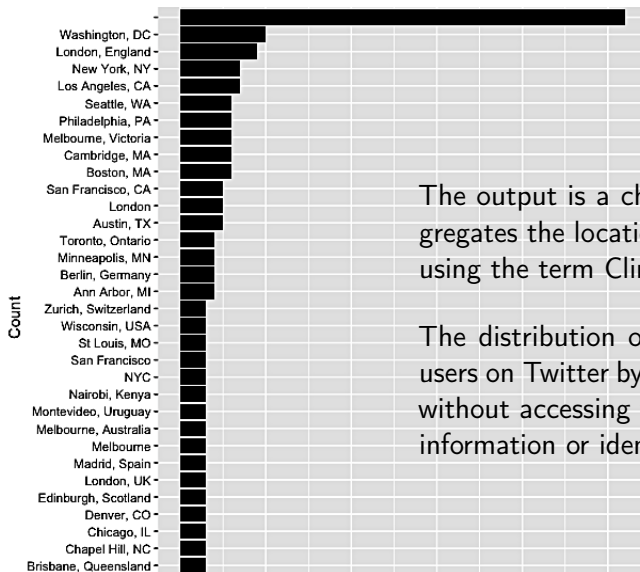
(1) Ethical considerations for collecting data from social media.

(2) How do I validate the hashtag public "ClimateStrike" on Twitter as a global conversation?

As a solution I've written an R script to scrape Twitter.



Where Twitter users are from - unique locations



The output is a chart that aggregates the location of people using the term ClimateStrike.

The distribution of the public users on Twitter by geolocation without accessing their private information or identity.



Social media data is a valuable research source for a critical discourse analysis.

Sensitivity considerations for collecting data from social media platforms.

Data reflects the political opinions of the contributors their identity may be discovered through an online search.



Discourse network Twitter for rtweet

Roslyn Walker, Macquarie University, Sydney

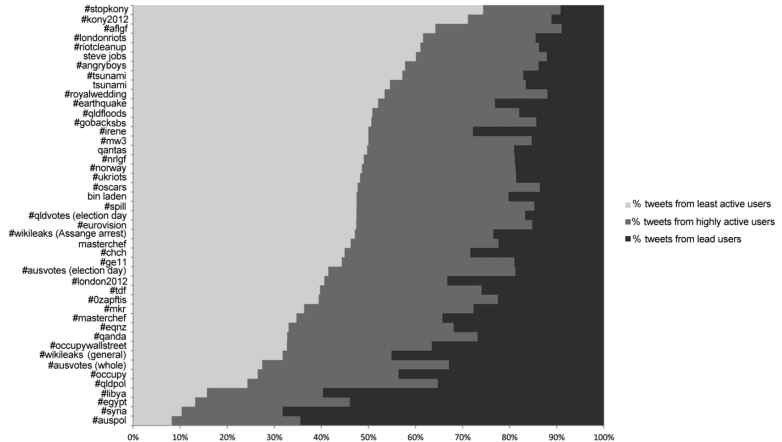


FIGURE 1 Relative contributions from the three user groups.

⇒ Image source: reproduced from Bruns and Steiglitz (2012: 171). Twitter activity patterns among user groups.



The screenshot displays the RStudio environment. The script editor on the left contains the following R code:

```
84 # note the words that are recognized as unique by R
85 a_list_of_words <- c("dog", "dog", "dog", "cat", "cat", ",")
86 unique(a_list_of_words)
87 # [1] "dog" "dog" "cat" ", "
88 # tidytext::unnest_tokens()
89 # remove punctuation, convert to lowercase, add id for each tweet!
90 clintestrike_tweets_clean <- clintestrike_tweets %>%
91   dplyr::select(stripped_text) %>%
92   unnest_tokens(word, stripped_text)
93 # plot the top 15 words -- notice any issues?
94 clintestrike_tweets_clean %>%
95   count(word, sort = TRUE) %>%
96   top_n(15) %>%
97   mutate(word = reorder(word, n)) %>%
98   ggplot(aes(x = word, y = n)) +
99   geom_col() +
100   xlab(NULL) +
101   coord_flip() +
102   labs(x = "Count",
103        y = "Unique words",
104        title = "Count of unique words found in tweets")
105 # load list of stop words - from the tidytext package
106 data("stop_words")
```

The console on the bottom left shows the following output:

```
~/Library/Mobile Documents/com~apple~CloudDocs/MQ Digital Humanities/Proof of concept/myrepo/myrepo/ <
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

The search results panel on the right shows the R logo and the text:

Search Results

The search string was "following object is masked_by"

No results found

My research uses rtweet in Rstudio to gather communicative actions within a hashtag community. This informs the information flow around an issue within and among communities.



I use Twitter as the digital technology for my research due to public status of tweets which are by default public. My study considers the affordance of digital technologies, such as Twitter, for civic protest for social change.

[more ...](#)



This proof of concept has three main goals:

- (1) send request to Twitter's stream APIs
- (2) retrieve data
- (3) format data into a structure

As a social media researcher, I wanted to access public social media postings as data, so that I could analyse the data collected under a hashtag public and identify discourse circulation patterns. As a social media researcher, I wanted to extract tweets based on a hashtag through metadata (place).



Quality assurance

The acceptance tests for this proof of concept includes whether the data retrieved is in an accessible format that enables the exporting, sorting, collecting, analysis, and archiving of the data output.

The acceptance criteria includes the use of automated tools to access and manipulate the data output.

These acceptances have been met by this proof of concept.



Limitations of the proof of concept

- (1) Access to data on Twitter requires a personal application programming interface (API) key. Source: <https://developer.twitter.com/en/apps>
- (2) Before using this software a user requires a secret pair for the OAuth flow.
- (3) The process of accessing Twitter changes frequently, so that software written to interact with the Twitter client requires frequent testing and revision.
- (4) None of the four private pieces of identity and authentication information should ever be committed to public source control, in order to protect your application and/or user account from compromise or misuse. Source: <https://twittercommunity.com/t/upcoming-changes-to-access-token-and-secret-management/130851>



References:

Bruns, A. (2018a) "Digital Public Spheres in Australia" in Digitizing Democracy, Routledge

Bruns, A. and S. Stieglitz (2012) Quantitative Approaches to Comparing Communication Patterns on Twitter, Journal of Technology in Human Services

Elster Hanson, J. (2019) 'Climate strike' named 2019 word of the year by Collins Dictionary' in The Guardian, 7 November 2019

Henderson-Sellers, A. (2010) "How seriously are we taking climate change? Monitoring climate change communication," Sydney: Sydney University Press.

Pew Center Research (2019) "A look at how people around the world view climate change" from the Pew Research Center's Spring 2018 Global Attitudes Survey.

Pico Presentation: <https://github.com/MQ-FOAR705/WalkerRoslynPICOpresentation>

Photographic images are the author's own ©Roslyn Walker 2019 Taken on location at the Global Strike, 20 September 2019 in Sydney, at the Domain.