

# Assignment 1

**BRACE YOURSELF**

**ASSIGNMENT IS COMING**

makeameme.org

[See Moodle](#)

# Review

Work with class partner to:

- List as many as possible the types of data that are freely available through Twitter API;
- List the types of data that are NOT available through the free-tier Twitter API

# Debugging time

[https://docs.google.com/document/d/141qWy-ucwr5\\_5pSKAlySSfIZDRfJrEyeb8SoBnSPsAM/edit](https://docs.google.com/document/d/141qWy-ucwr5_5pSKAlySSfIZDRfJrEyeb8SoBnSPsAM/edit)

# The Types of Free and Available Data

- Recent tweets that match given keywords/hashtags
- A user's recent 3,200 timeline tweets
- A user's followers and friends
- A list of users matching given hashtags/keywords on their Twitter bio

# The Types of Data NOT Freely Available

- Historical tweets
- The complete Twitter streams

# R Function Explained

In R, a *function* performs a task based on inputs and predefined logics.

```
mytoken <- create_token(  
  app = "", #app name here  
  consumer_key = "", #consumer key here  
  consumer_secret = "", #consumer secret here  
  access_token = "", #access token here  
  access_secret = "" ) #access secret here
```

This is a function. It's called `create_token()`. It authenticates Twitter API based on a number of parameters.

The parameters...

# R Function Explained

In the examples below, identify R functions and parameters for the function.

```
tweets1 <- search_tweets("#polarvortex", n = 100, token=mytoken)
```

```
save_as_csv(mentions, "trump's_mentions.csv")
```

# R Cheat Sheet

A list of R functions commonly used in collecting Twitter data. You will use them in tutorials and practice scripts. Do you recall using any of them?

- `create_token()`
- `search_tweets()` or `stream_tweets()`
- `get_timelines()`
- `lookup_users()`
- `get_followers()`
- `get_friends()`
- `search_users()`



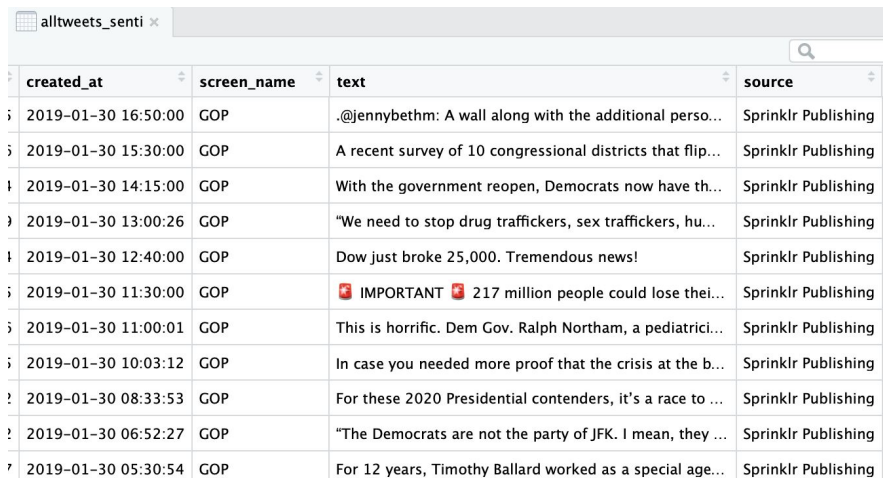
## Mini hackathon (15 mins)

Work with your class partner to:

- Get tweets based on a keyword/hashtag or get tweets from a user's timeline;
- Find the most retweeted tweet in your dataset;
- Each team will compete for finding the most viral tweet.

# Explore the Metadata in Your Dataset

- Based on the tweets your team has collected, list all available metadata (perhaps you should review the definition of metadata).
- Hint: the function `colnames()` can give you a list of columns in your data frame.



The image shows a screenshot of a data frame viewer for a dataset named 'alltweets\_senti'. The table has four columns: 'created\_at', 'screen\_name', 'text', and 'source'. The data rows show tweets from January 30, 2019, with various text and sources.

created_at	screen_name	text	source
2019-01-30 16:50:00	GOP	.@jennybethm: A wall along with the additional perso...	Sprinklr Publishing
2019-01-30 15:30:00	GOP	A recent survey of 10 congressional districts that flip...	Sprinklr Publishing
2019-01-30 14:15:00	GOP	With the government reopen, Democrats now have th...	Sprinklr Publishing
2019-01-30 13:00:26	GOP	"We need to stop drug traffickers, sex traffickers, hu...	Sprinklr Publishing
2019-01-30 12:40:00	GOP	Dow just broke 25,000. Tremendous news!	Sprinklr Publishing
2019-01-30 11:30:00	GOP	🚨 IMPORTANT 🚨 217 million people could lose thei...	Sprinklr Publishing
2019-01-30 11:00:01	GOP	This is horrific. Dem Gov. Ralph Northam, a pediatri...	Sprinklr Publishing
2019-01-30 10:03:12	GOP	In case you needed more proof that the crisis at the b...	Sprinklr Publishing
2019-01-30 08:33:53	GOP	For these 2020 Presidential contenders, it's a race to ...	Sprinklr Publishing
2019-01-30 06:52:27	GOP	"The Democrats are not the party of JFK. I mean, they ...	Sprinklr Publishing
2019-01-30 05:30:54	GOP	For 12 years, Timothy Ballard worked as a special age...	Sprinklr Publishing

# Brainstorming

- Imagine what insights can be gained from the available Twitter metadata. List as many possibilities as possible (use your imagination).

# Brainstorming

- Based on the type of metadata available, can you propose something that uses the data for social good?
- Can you also think of ways in which having access to such data may result in unintended (or intended) bad consequence?

Good examples:

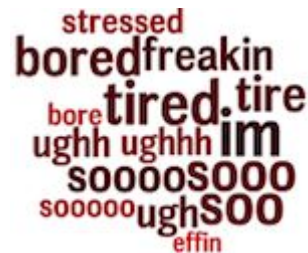
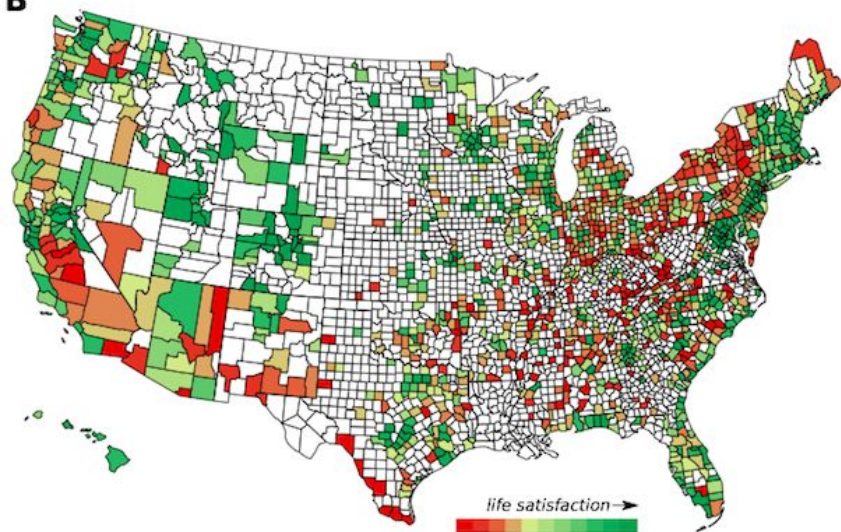
- [Mapping Twitter data for disaster response](#)

# Brainstorming

A good but debatable example

- [Use tweets to predict the wellbeing of a community](#)

B



# Brainstorming

- Work with your class partner to review the metadata in your dataset. Mark those that you think are private. Explain why they are private information.

# In-class practice

W3\_TW2.R

# Required tutorials for THIS WEEK

An interactive tutorial  
for COMM 497DB

Weilai Wayne Xu

Libraries/packages

Data frames

Connecting to the Twitter API

Collect tweets by keywords/hashtags

Collect Twitter user timeline

Collect Twitter user info

Bonus: Collect YouTube Data

Make Wordclouds

Visualizing virality

Predict Ideology (in progress)

## Using R for Digital Behavior Analyt

### Libraries/packages

**What is a library/package?** Think of R as an operating system (e.g., iOS, V system. Each library is designed to accomplish specific tasks. For example, the semester—is for visualizing data, and the library *rtweet* is used for collect

Use **install.packages()** to install libraries. Use **library()**, or **require()** to load

Next, we will install a fun library called *cowsay*.

Code

Start Over

```
1 # install.packages("cowsay")
2 #make sure the library name is wrapped by quotation.
3
4 library(cowsay) #load the library, alternatively, you can us
```

This tutorial is hosted on a cloud server, running the above code won't have and run it on your local machine. Keep an eye on what is happening in the C

Let's have some fun with *cowsay*.

Run the code and see what happens.