

Review

Get in a group of three and work on the following questions

- What is a R library/package?
- What is a R data frame?
- Share with your group members any error/difficulty/frustration you've faced when working on tutorials

Review

Code

↻ Start Over

▶ Run Code

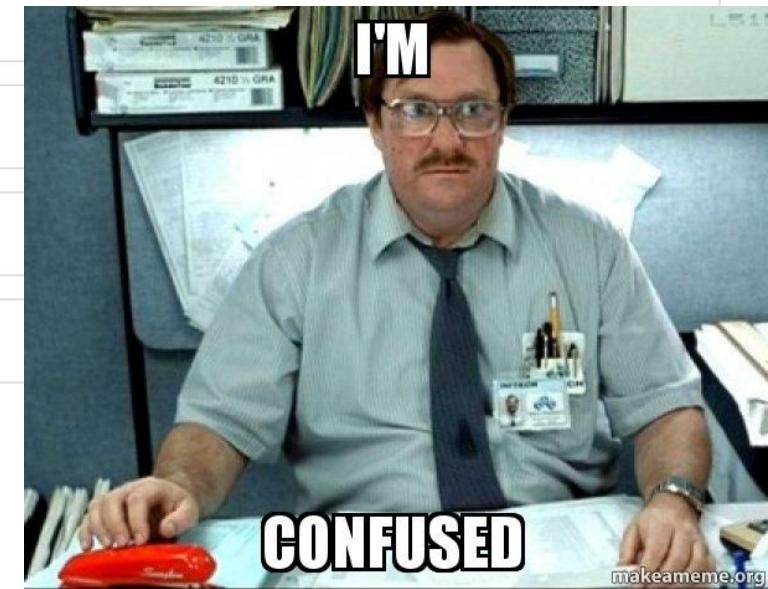
```
1 library(rtweet)
2 t <- stream_tweets("news", include_rts = FALSE, timeout = 6)
3 t
```

Streaming tweets for 6 seconds...

The stream disconnected prematurely. Reconnecting...

Reconnecting again...

NULL



Review

Why did I get this error?

Blame the “rate limiting” imposed by Twitter! Twitter caps how many data requests you can make within a 15 minute window.



**SEE ONE,
DO ONE,
TEACH ONE.**

A graphic element consisting of a horizontal line with a small diamond shape in the center.

Application Programming Interfaces (API)

Weiai Xu (Wayne), PhD

Assistant Professor

Department of Communication, UMass-Amherst

Email: weiaixu@umass.edu

curiositybits.cc

Activity

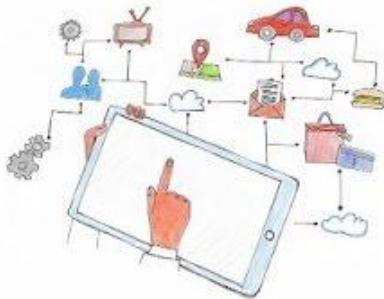
Get in a group of three

Do web searches to find out:

What Exactly is an API?

API according to YouTube videos

DEMYSTIFYING THE API



https://www.youtube.com/results?search_query=what+is+api

What is API anyway???

An API is like a doorman leading you to an internet platform's data treasures.

To access the data treasures, you need to **“swipe your key card”** and **“speak to”** the doorman in a language that he/she understands.



But first, let's talk about two conventional ways of collecting digital data

1. Web crawling (or called *screen-scraping*)
2. API

Use W2_API.R (available on Moodle)

Workflow in web crawling

1. Web crawling

Voting members by state [edit]

District	Representative	Party	Prior experience	Education	Assumed office	Residence	Born
Alabama 1	Bradley Byrne	Republican	Alabama Senate Alabama State Board of Education	Duke University (BA) University of Alabama (JD)	2014*	Fairhope	1955
Alabama 2	Martha Roby		Montgomery City Council	New York University (BM) Samford University (JD)			
Alabama 3	Mike Rogers		Calhoun County Commissioner Alabama House of Representatives	Jacksonville State University (BA, MPA) Birmingham School of Law (JD)	1 Alabama 1	Bradley Byrne	
Alabama 4	Robert Aderholt		Haleyville Municipal Judge	University of North Alabama Birmingham-Southern College (BA) Samford	2 Alabama 2	Martha Roby	



```
library(rvest)
library(tidyverse)
h <- read_html("https://en.wikipedia.org/wiki/Current_members_of_the_United_States_House_of_Representatives")
reps <- h %>%
  html_node("table#votingmembers") %>%
  html_table(fill = TRUE)
```

District	Representative	Party	Prior experience	Education	Assumed office
1 Alabama 1	Bradley Byrne	Republican	Alabama Senate Alabama State Board of Education	Duke University (BA) University of Alabama (JD)	2014*
2 Alabama 2	Martha Roby	Republican	Montgomery City Council	New York University (BM) Samford University (JD)	2011
3 Alabama 3	Mike Rogers	Republican	Calhoun County Commissioner Alabama House of Representatives	Jacksonville State University (BA, MPA) Birmingham School of Law (JD)	2003
4 Alabama 4	Robert Aderholt	Republican	Haleyville Municipal Judge	Haleyville Municipal Judge	University of North Alabama Birmingham-Southern College (BA) Samford
5 Alabama 5	Mo Brooks	Republican	Alabama House of Representatives	Madison County Co...	Duke University (BA) University of Alabama (JD)
6 Alabama 6	Gary Palmer	Republican	University of North Alabama Birmingham-Southern College (BA) Samford	Policy analyst	2011
7 Alabama 7	Terri Sewell	Democratic		Attorney	University of Alabama (BS)
8 Alaska at large	Don Young	Republican		Alaska Senate Ship captain Mayor of Fort Yukon, Alaska	2015
9 Arizona 1	Tom O'Halleran	Democratic		Arizona Senate	Princeton University (BA) St Hilda's College, Oxford (M...
10 Arizona 2	Ann Kirkpatrick	Democratic		U.S. House, Arizona House of Representatives	2011
11 Arizona 3	Raúl Grijalva	Democratic		Pima County Board of Supervisors	Brigham Young University (BA) University of Arizona (BA)
12 Arizona 4	Paul Gosar	Republican		President of the Northern Arizona Dental Society	2003
13 Arizona 5	Andy Biggs	Republican		Arizona Senate	Creighton University (BS, DDS)
14 Arizona 6	David Schweikert	Republican		Arizona House of Representatives	2017
15 Arizona 7	Ruben Gallego	Democratic		Arizona Board of Ed...	Brigham Young University (BA) University of Arizona (J...
16 Arizona 8	Debbie Lesko	Republican		Arizona House of Representatives	2011
17 Arizona 9	Greg Stanton	Democratic		Arizona Senate Presi...	Arizona State University, Tempe (BS, MBA)
18 Arkansas 1	Rick Crawford	Republican		Mayor of Phoenix	2015
19 Arkansas 2	French Hill	Republican		Broadcaster, businessman	Marquette University (BA) University of Michigan (JD)
				Businessman	Arkansas State University (BS)
					2011
					Vanderbilt University (BS)
					2018*

Workflow in web crawling

Partisan mix of the House by state [edit]

Partisan mix of the House by state

[show]

Voting members by state [edit]

District	Representative	Party	Prior experience	Education
Alabama 1	 Bradley Byrne		Alabama Senate Alabama State Board of Education	Duke University of Alabama (JL)
Alabama 2	 Martha Roby		Montgomery City Council	New York University (I) Samford University
Alabama 3	 Mike Rogers		Calhoun County Commissioner Alabama House	Jacksonville State University (I) MPA)

```
▶ <nz>...</nz>
▶ <div class="mw-collapsible mw-collapsed mw-made-collapsible" style="box-sizing:border-box;width:100%;font-size:95%;padding:4px;border:none;">...
▶ <h2>...</h2>
▼ <table class="wikitable sortable jquery-tablesorter" id="votingmembers">
  ▶ <thead>...</thead>
  ▶ <tbody>
    ▶ <tr>
      ▶ <td>...</td>
      ▶ <td nowrap="nowrap">...</td>
      <td rowspan="6" style="background-color:#E81B23">
      ▶ <td rowspan="6">Republican
      ▶ <td>...</td> == $0
      ▶ <td>
        <a href="/wiki/Duke_University" title="Duke University">Duke University</a>
        ▶ <span style="font-size:85%;">...</span>
      <br>
        <a href="/wiki/University_of_Alabama" title="University of Alabama">University of Alabama</a>
        ▶ <span style="font-size:85%;">...</span>
      </td>
      ▶ <td>2014*</td>
      ▶ <td>...</td>
      <td>1955
      ▶ </tr>
    ▶ <tr>...</tr>
```

Problems with web crawling

1. Based on extracting information from a webpage's HTML or XML codes. But each page has different formats and layouts. Thus, you need to tailor codes for each site;
2. Web crawling is like a **bot** automatically visiting and downloading data from websites. It could violate the “Terms of Service” of some websites. Most social media platforms simply do not allow web crawling.

4. PROHIBITED USE OF THE SERVICES

You may not access or use, or attempt to access or use, the Services to take any action that could harm us or a third party. You may not use the Services in violation of applicable laws or in violation of our or any third party's intellectual property or other proprietary or legal rights. You further agree that you shall not attempt (or encourage or support anyone else's attempt) to circumvent, reverse engineer, decrypt, or otherwise alter or interfere with the Services, or any content thereof, or make any unauthorized use thereof. Without NYT's prior written consent, you shall not:

(i) access any part of the Services, Content, data or information you do not have permission or authorization to access or for which NYT has revoked your access;

(ii) use robots, spiders, scripts, service, software or any manual or automatic device, tool, or process designed to data mine or scrape the Content, data or information from the Services, or otherwise access or collect the Content, data or information from the Services using automated means;

<https://help.nytimes.com/hc/en-us/articles/115014893428-Terms-of-service>

Try web crawling in R

```
library(rvest)
library(tidyverse)
h <- read_html("https://en.wikipedia.org/wiki/Current_member

reps <- h %>%
  html_node("table#votingmembers") %>%
  html_table(fill = TRUE)
```

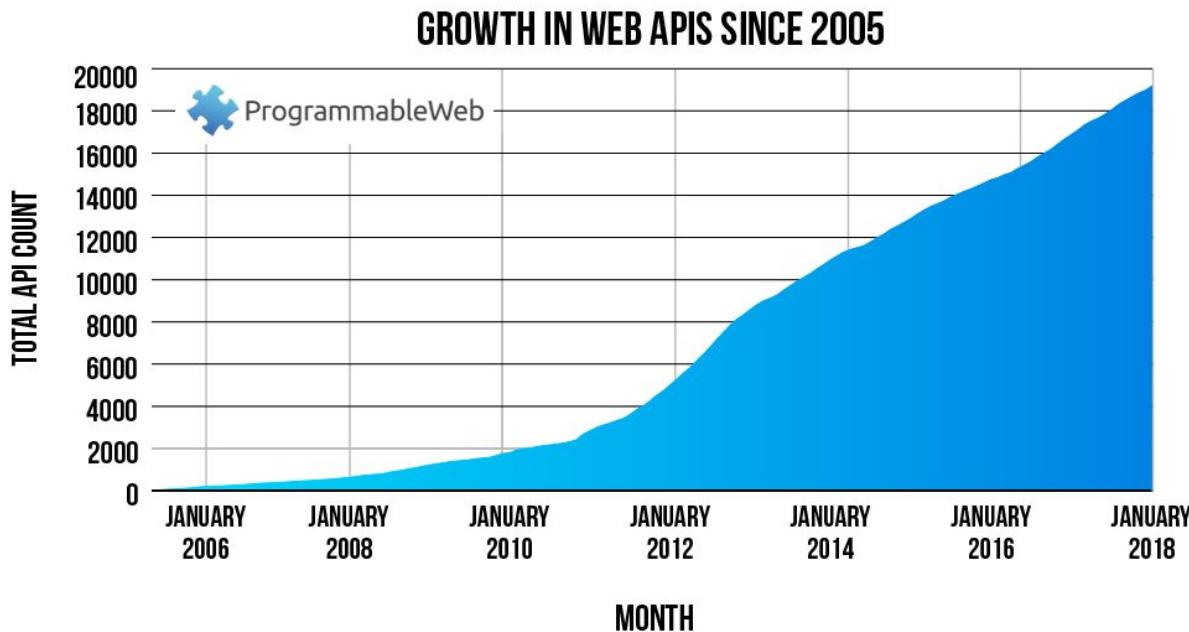
In **W2_API.R**

You would need to install two new libraries: *rvest* and *tidyverse*

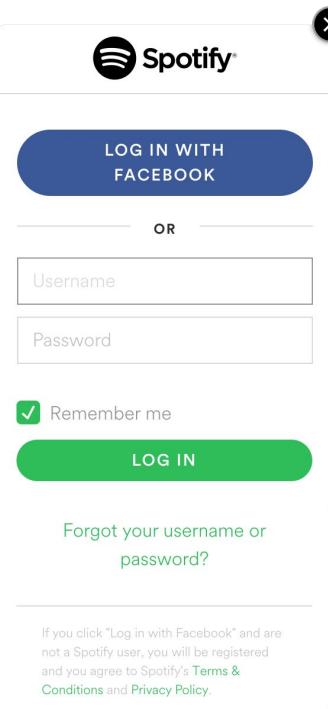
API

There are over 20,728 APIs out there. Most web platforms that you are familiar with have APIs for data sharing.

However, APIs also become increasingly restrictive.



APIs for cross-platform data sharing



Use case 1: You sign up Spotify using your Facebook account. Spotify connects to the Facebook API to retrieve your account information.

But...

Data-sharing between platforms could cause privacy concerns

Three screenshots illustrating data sharing. The first is a screenshot of a Facebook sidebar menu. It shows links like News Feed, Messenger, Watch, Marketplace, and Shortcuts. The second is a screenshot of an Amazon sponsored ad for a NUU Mobile G3 smartphone. The third is a screenshot of an Amazon product listing for a Moto G6 smartphone.

If you click "Log in with Facebook" and are not a Spotify user, you will be registered and you agree to Spotify's [Terms & Conditions](#) and [Privacy Policy](#).

APIs for data collection

Use case 2: Collect structured data from Twitter for analytics

API credentials

You need to “swipe your key card” to access API. For Twitter API, you need the following:

App name, API key, API secret key, Access token, Access token secret

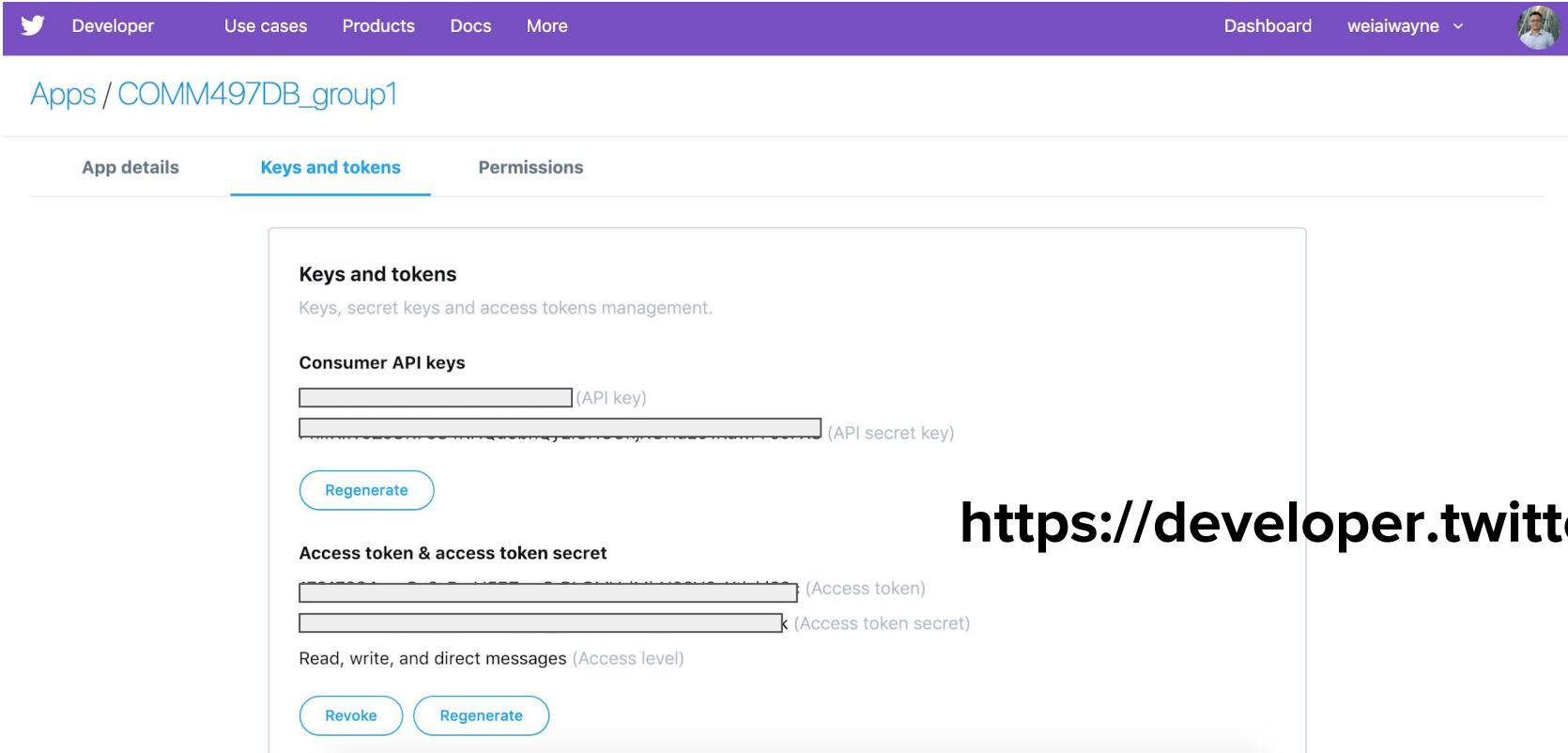
Before making calls to Twitter’s API, you need to authenticate ourselves vis-a-vis Twitter’s API. This is done through the **create_token()** function in **rtweet** library (see the code example in the next slide).

```
#replace the following API credentials with the one posted on Moodle.  
mytoken <- create_token(  
  app = "APP1", #app name here  
  consumer_key = "DSD62iGWw16nMwaCSLkzfSQA", #consumer key here  
  consumer_secret = "p62iGWw16nMwaCSLkzfSQA", #consumer secret here  
  access_token = "153474365-XuWYfm1E423Ew6yuUM6Jfm7GMRHWJXzclWNPGCFmM", #access token here  
  access_secret = "tHq0Hq0xAqhaXHWlkXQ76HBQ7NVIX0rwvGRiH5cnsNE") #access secret here
```

This is JUST A CODE DEMO!

The API credentials shown above are NOT valid. Refer to the complete code in W2_API.R

Where do we get the API credentials



The screenshot shows the Twitter Developer Keys and tokens page. At the top, there's a purple navigation bar with links for Developer, Use cases, Products, Docs, More, Dashboard, and a user profile for 'weiaiwayne'. Below the navigation bar, the URL 'Apps / COMM497DB_group1' is visible. The main content area has tabs for App details, Keys and tokens (which is selected), and Permissions. The 'Keys and tokens' section contains sections for Consumer API keys and Access token & access token secret, each with fields for API key, API secret key, Access token, and Access token secret, along with Regenerate and Revoke buttons.

Keys and tokens

Keys, secret keys and access tokens management.

Consumer API keys

[REDACTED] (API key)
[REDACTED] (API secret key)

Regenerate

Access token & access token secret

[REDACTED] (Access token)
[REDACTED] (Access token secret)

Read, write, and direct messages (Access level)

Revoke Regenerate

<https://developer.twitter.com>

APIs for data collection

Check which group you are assigned to on Moodle.

API credentials are provided in group description.

How to obtain the Twitter API credentials

Prior to August, 2018, practically any Twitter user could create a Twitter app to obtain API credentials.

Due to the increasing public scrutiny over social media companies' practices in data protection, Twitter announced a big API update in 2018 that allows only **Twitter developer accounts** to obtain API credentials.

https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform.html

How to obtain the Twitter API credentials

You need to apply for a developer account. You will need to provide a cell-phone number and explain to Twitter what you intend to do with the app. It takes weeks or even months for Twitter to vet your application.

https://cbail.github.io/textasdata/apis/rmarkdown/Application_Programming_interfaces.html



Application under review.

Thanks! We've received your application and are reviewing it. We'll be in touch soon.

We review applications to ensure compliance with our Terms of Service and Developer policies. [Learn more.](#)

You'll receive an email when the review is complete. While you wait, check out our [documentation](#), explore our [tutorials](#), or check out our [community forums](#).

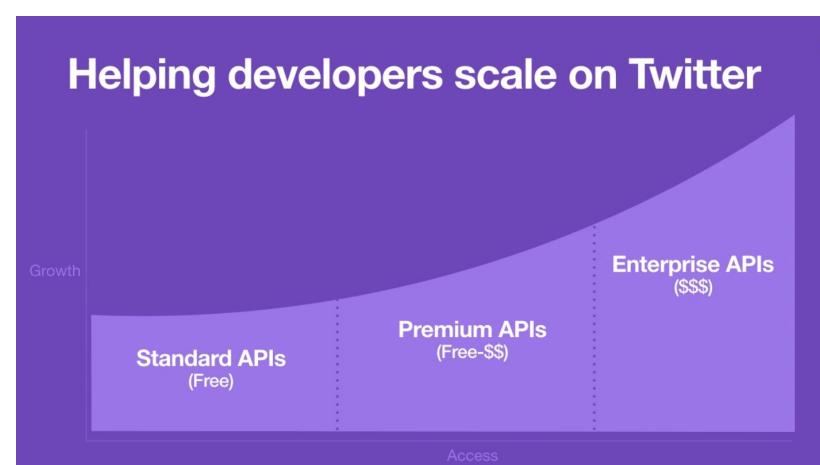
Rate limits in APIs

Internet is no longer the wild west when it comes to harvesting user data. Various web platforms impose strict rate limiting, restricting how much information an API user can collect within a period of time.

Rate limits for searching tweets:

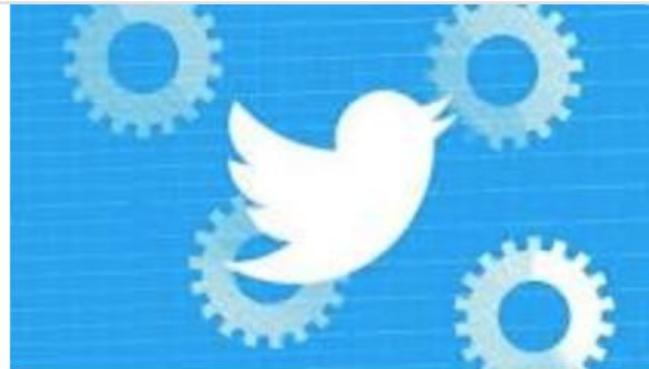
180 API calls within a 15-minute window
(notice that this rule may change over time).

Want less restriction, apply for the expensive Twitter Premium APIs!



Data monetization

The pricing for the premium APIs ranges from \$149/month to \$2,499/month, based on the level of access needed. The first premium offering, the Search Tweets API, is launching today into public beta. This will give developers the ability to access the past 30 days of Twitter data. Nov 14, 2017



Twitter launches lower-cost subscription access to its data through ...
<https://techcrunch.com/.../twitter-launches-lower-cost-subscription-access-to-its-data-thro...>

Other limits in APIs

Historical tweets are NOT available through the free-tier Twitter API.

Want access? **Buy** data from Twitter or third-party data vendors (e.g., Gnip).

See the estimate provided by Sifter (a company in Amherst MA, no longer in business after the recent API change)

Hi Weiai Wayne,

The estimate has completed for Job: 20180323213156-7034

Rule Text: #netneutrality

Start Date: 12/01/2017

End Date: 12/20/2017

Estimated Activities: 5,881,000

The total cost of accepting this job is \$4917.50.

API will change like shifting sands

Web platforms are making APIs more restrictive for average developers/users. The access to and the power of data is now concentrated on platforms that own user data.

Twitter's recent API change has made some data analytics companies obsolete.



[home](#) [testimonials](#) [faq](#) [discovertext](#) [contact](#)

Unfortunately, this site has been decommissioned as of Sept. 30, 2018.

Thanks to our 6,970 users who created 16,128 free estimates from the complete, undeleted history of Twitter between 1/14/2014 and 9/29/2018. Please contact Twitter for approval of future academic or commercial use cases. If you can get an approved use case from Twitter, we can still help you work with the data inside [DiscoverText](#). [@DiscoverText](#) remains open and is still the **top-ranked text analysis platform on the Internet**.

All paid jobs prior to the decommissioning will still be honored.

API will change like shifting sands

Some big API changes that have occurred to social media platforms



Previously, every Twitter user could access Twitter's API for free, with some restrictions.

Open only to Twitter developer accounts.



Previously, you could download a public Facebook page's posts and comments for free.

No access to public Facebook page data



Deen Freelon

@dfreelon

Following

So Facebook has shuttered API access without direct approval:

[newsroom.fb.com/news/2018/04/r ...](https://newsroom.fb.com/news/2018/04/restricting-data-access/)

Party's over, folks--time to start thinking about what happens if/when Twitter does the same.

“
These changes will better protect people’s information while still enabling developers to create useful experiences.”

An Update on Our Plans to Restrict Data Access on Facebook

Two weeks ago we promised to take a hard look at the information apps can use when you connect them to Facebook as well as other data practices. Today, we

newsroom.fb.com

10:04 PM - 5 Apr 2018

<https://newsroom.fb.com/news/2018/04/restricting-data-access/>

April 4, 2018

An Update on Our Plans to Restrict Data Access on Facebook

“

We believe these changes will better protect people’s information while still enabling developers to create useful experiences.

”

Pages API: Until today, any app could use the Pages API to read posts or comments from any Page. This let developers create tools for Page owners to help them do things like schedule posts and reply to comments or messages. But it also let apps access more data than necessary. We want to make sure Page information is only available to apps providing useful services to our community. So starting today, all future access to the Pages API will need to be approved by Facebook.

Food for thoughts

- Why do web platforms provide APIs?
- Why do web platforms make APIs more restrictive?
- Having reviewed the current API rules and recent API changes, what's your take on the issue of equality in data protection and data access?

Try it yourself

Use **W_API.R** (it is on Moodle!)

W_API.R has two parts: one for doing the traditional web crawling (which you can safely skip), and the other for connecting to Twitter's API. Make sure you can connect to the API using the credential I've provided.

7-9 students will share the same credential. In order not to hit the rate limit, set a small n .

Required tutorials for this week

An interactive tutorial
for COMM 497DB

Weiwei Wayne Xu

Libraries/packages

Data frames

Connecting to the Twitter API

Collect tweets by keywords/hashtags

Collect Twitter user timeline

Collect Twitter user info

Make Wordclouds

Predict Ideology (in progress)

Collect Twitter user info

Collecting user information? That sounds creepy!

Not at all. We will conduct the data collection in strict compliance with Twitter's developer terms. In fact, just like on collecting tweets, Twitter makes it very limited as to what kind of user profile data are available through its API.

What can you do with the Twitter user data?

After running the code in this part of the tutorial, you will end up with a data frame containing a bunch of screen names and profile bios. You might ask: what can I do with it? You will be surprised by how much insights we can draw by just looking at these profiles. For example, we can use artificial intelligence to predict a user's ideology. We will, of course, save the predicted ideologies in a separate column.

Get followers and friends

Running the code below, we can get Nassim Nicholas Taleb's (@nntaleb) followers and friends. Nassim Nicholas Taleb is a favorite author of mine. He wrote the famous [*The Black Swan: The Impact of the Highly Improbable: With a new section on Fragility*](#). Interestingly, Taleb served as a [UMass-Amherst faculty](#) from January 2005 to January 2006.

Code



Start Over

Start Over

Wanna grab YouTube metadata? Try Google's API

≡ Google Cloud Platform



New Project



You have 21 projects remaining in your quota. Request an increase or delete projects.

[Learn more](#)

[MANAGE QUOTAS](#)

Project Name *

COMM497DB



Project ID: comm497db. It cannot be changed later. [EDIT](#)

Location *

No organization

[BROWSE](#)

Parent organization or folder

[CREATE](#)

[CANCEL](#)

<https://console.cloud.google.com/apis/api>

Wanna grab YouTube metadata? Try Google's API

The screenshot shows the Google Cloud Platform API Library interface. At the top, there is a blue header bar with the text "Google Cloud Platform" and "COMM497DB". A search bar contains the text "YOUTUBE". To the right of the search bar are several icons: a magnifying glass, a question mark, a person icon with a "1", and a vertical ellipsis. Below the header, a search bar shows the term "YOUTUBE". On the left, a sidebar has a "Search" button and a back arrow. The main content area displays "4 results" for "YOUTUBE". Each result card includes a thumbnail (a red play button), the API name, the provider (Google), and a brief description. The first result is "YouTube Data API v3" by Google, which provides access to YouTube data like videos and playlists. The second result is "YouTube Analytics API" by Google, which retrieves YouTube Analy data. The third result is "YouTube Ads Reach API" by Google. The fourth result is "YouTube Reporting API" by Google. A large circular thumbnail for "YouTube Data API v3" is shown at the bottom right. Below the cards, a blue "ENABLE" button and a white "TRY THIS API" button are visible. A blue arrow points from the text "API Library" to the "YouTube Data API v3" card.

Google Cloud Platform COMM497DB

YOUTUBE

4 results

YouTube Data API v3

YouTube Analytics API

YouTube Ads Reach API

YouTube Reporting API

API Library

YouTube Data API v3

ENABLE TRY THIS API

Enable all YouTube APIs. Also enable Freebase API

Wanna grab YouTube metadata? Try Google's API

The screenshot shows two views of the Google Cloud Platform API Library. On the left, a modal window displays the 'YouTube Data API v3' page, which includes a large YouTube logo icon, the service name 'Google', a brief description, and two buttons: 'MANAGE' (highlighted with a red box) and 'TRY THIS API'. On the right, the main 'Overview' page for the YouTube Data API v3 is shown. It features a sidebar with 'APIs & Services' and 'YouTube Data API v3'. The main content area includes a note about needing credentials, a 'CREATE CREDENTIALS' button (also highlighted with a red box), a 'Details' section with service information, and a 'Traffic by response code' chart.

YouTube Data API v3

Google

The YouTube Data API v3 is an API that provides access to YouTube data, such as videos, playlists,...

MANAGE TRY THIS API

APIs & Services

YouTube Data API v3

Overview DISABLE API PROVIDE FEEDBACK

To use this API, you may need credentials. Click 'Create credentials' to get started.

CREATE CREDENTIALS

Details

Name: YouTube Data API v3
By: Google
Service name: youtube.googleapis.com

Overview

The YouTube Data API v3 is an API that provides access to YouTube data, such as videos, playlists, and channels.

Activation status: Enabled

Traffic by response code

Request/sec (2 hr average)

Response Code	Request/sec (2 hr average)
1.0/s	1.0/s
0.8/s	0.8/s
0.6/s	0.6/s
0.4/s	0.4/s
0.2/s	0.2/s
0.0/s	0.0/s

No data is available for the selected time frame.

2019 2019 2019 2019

Enable all YouTube APIs. Also enable Firebase API

Wanna grab YouTube metadata? Try Google's API

Credentials

[Credentials](#) [OAuth consent screen](#) [Domain verification](#)

The screenshot shows the 'Credentials' section of the Google Cloud Platform console. It lists four options: 'API key', 'OAuth client ID', 'Service account key', and 'Help me choose'. The 'OAuth client ID' option is highlighted with a red box. At the bottom, there is a blue button labeled 'Create credentials ▾'.

- API key
Identifies your project using a simple API key to check quota and access
- OAuth client ID**
Requests user consent so your app can access the user's data
- Service account key
Enables server-to-server, app-level authentication using robot accounts
- Help me choose
Asks a few questions to help you decide which type of credential to use

Create credentials ▾

Wanna grab YouTube metadata? Try Google's API

[←](#) Create OAuth client ID

For applications that use the OAuth 2.0 protocol to call Google APIs, you can use an OAuth 2.0 client ID to generate an access token. The token contains a unique identifier. See [Setting up OAuth 2.0](#) for more information.

Application type

- Web application
- Android [Learn more](#)
- Chrome App [Learn more](#)
- iOS [Learn more](#)
- Other

Name [?](#)

app2

[Create](#)

[Cancel](#)

Wanna grab YouTube metadata? Try Google's API

Credentials

Credentials OAuth consent screen Domain verification

Create credentials

Create credentials t

OAuth 2.0 client

Name

app2

tuber1

OAuth client

The client ID and secret can always be accessed from Credentials in APIs & Services

Info OAuth is limited to 100 [sensitive scope logins](#) until the [OAuth consent screen](#) is published. This may require a verification process that can take several days.

Here is your client ID

[REDACTED]

Here is your client secret

[REDACTED]

OK

This screenshot shows the 'Credentials' section of the Google Cloud Platform (GCP) console. Under the 'OAuth 2.0 client' tab, a new client has been created with the name 'app2'. The 'Client ID' and 'Client secret' fields are displayed as redacted text with a 'Copy' button next to each. A tooltip provides information about the limit of 100 sensitive scope logins and the need for an OAuth consent screen, which requires verification and can take several days. The 'OK' button at the bottom right of the modal indicates the user is finished with this step.

Wanna grab YouTube metadata? Try Google's API

```
library(tuber)

#connect to YouTube's API. More at https://github.com/soodoku/tuber
yt_oauth("enter Client ID here", "enter Client secret here", token = '')
```

Wanna grab YouTube metadata? Try Google's API

The screenshot shows a web browser window with the following details:

- Header:** A "Sign in with Google" button is visible on the left.
- Address Bar:** The URL is `localhost:1410/?state=6zEo75EvSc&code=4/4ABz6TzHb6k4`.
- Toolbar:** Standard browser controls (back, forward, refresh) are at the top.
- Content Area:**
 - A message says "Choose an account to continue to tuber".
 - An "Authentication complete. Please close this page and return to R." message is displayed.
 - A user profile for "Weiai Xu" (weiai.wayne.xu@gmail.com) is shown.
 - A link to "Use another account" is present.
- Bottom Overlay:** A semi-transparent box contains the text:

Waiting for authentication in browser...
Press Esc/Ctrl + C to abort
Authentication complete.
>

Wanna grab YouTube metadata? Try Google's API

Video id

https://www.youtube.com/watch?v=0JMkzakXgIY

Apps Google DuckDuckGo Google Calendar My Drive - Google... Inbox (6) - weiai.... Google

YouTube Search

The Washington Post

SUBSCRIBE

0:00 / 4:11

CC HD

Try it yourself

Use **W_API.R** (it is on Moodle!).

The third part of the script is for using YouTube's API.