day_three

May 11, 2016

1 Creating data with web APIs

Most people who think they want to do web scraping actually want to pull data down from site-supplied APIs. Using an API is better in almost every way, and really the only reason to scrape data is if:

1. The website was constructed in the 90s and does not have an API; or,
2. You are doing something illegal

If LiveJournal has an API, the website you are interested in probably does too.

1.1 What is an API?

API is shorthand for **A**pplication **P**rogramming **I**nterface, which is in turn computer-ese for a middleman.

Think about it this way. You have a bunch of things on your computer that you want other people to be able to look at. Some of them are static documents, some of them call programs in real time, and some of them are programs themselves.

Solution 1 You publish login credentials on the internet, and let anyone log into your computer
Problems:

1. People will need to know how each document and program works to be able to access their data

2. You don't want the world looking at your browser history

Solution 2 You paste everything into HTML and publish it on the internet
Problems:

1. This can be information overload

2. Making things dynamic can be tricky

Solution 3 You create a set of methods to act as an intermediary between the people you want to help and the things you want them to have access to.
Why this is the best solution:

1. People only access what you want them to have, in the way that you want them to have it

2. People use one language to get the things they want

Why this is still not Panglossian:

1. You will have to explain to people how to use your middleman

1.2 Twitter's API

Twitter has an API - mostly written for third-party apps - that is comparatively straightforward and gives you access to *nearly* all of the information that Twitter has about its users, including:

1. User histories

2. User (and tweet) location

3. User language

4. Tweet popularity

5. Tweet spread

6. Conversation chains

Also, Twitter returns data to you in json, or **J**ava **S**cript **O**bject **N**otation. This is a very common format for passing data around http connections for browsers and servers, so many APIs return it as a datatype as well (instead of using something like xml or plain text).

Luckily, json converts into native Python data structures. Specifically, every json object you get from Twitter will be a combination of nested `dicts` and `lists`, which you learned about yesterday. This makes Twitter a lot easier to manipulate in Python than html objects, for example.

Here's what a tweet looks like:

```
In [1]: import json

        with open('../data/02_tweet.json','r') as f:
            a_tweet = json.loads(f.read())
```

We can take a quick look at the structure by pretty printing it:

```
In [2]: from pprint import pprint

        pprint(a_tweet)
```

```
{'contributors': None,
 'coordinates': None,
 'created_at': 'Thu Apr 02 06:09:39 +0000 2015',
 'entities': {'hashtags': [], 'symbols': [], 'urls': [], 'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 583511591334719488,
 'id_str': '583511591334719488',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'lang': 'ht',
 'place': None,
 'retweet_count': 0,
 'retweeted': False,
 'source': '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
 'text': '.IPA rettiwT eht tuoba nraeL',
 'truncated': False,
 'user': {'contributors_enabled': False,
```

```
'created_at': 'Thu Apr 02 05:54:25 +0000 2015',
'default_profile': True,
'default_profile_image': False,
'description': '',
'entities': {'description': {'urls': []}},
'favourites_count': 0,
'follow_request_sent': False,
'followers_count': 0,
'following': False,
'friends_count': 0,
'geo_enabled': False,
'id': 3129088320,
'id_str': '3129088320',
'is_translation_enabled': False,
'is_translator': False,
'lang': 'en',
'listed_count': 0,
'location': '',
'name': 'Yelekreb Bald',
'notifications': False,
'profile_background_color': 'C0DEED',
'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
'profile_background_tile': False,
'profile_image_url': 'http://pbs.twimg.com/profile_images/583509317476712449/mkd8KGeu_normal.jp
'profile_image_url_https': 'https://pbs.twimg.com/profile_images/583509317476712449/mkd8KGeu_no
'profile_link_color': '0084B4',
'profile_location': None,
'profile_sidebar_border_color': 'C0DEED',
'profile_sidebar_fill_color': 'DDEEF6',
'profile_text_color': '333333',
'profile_use_background_image': True,
'protected': False,
'screen_name': 'tob_pohskrow',
'statuses_count': 1,
'time_zone': None,
'url': None,
'utc_offset': None,
'verified': False}}
```

Time for a challenge!   Let's see how much you remember about lists and dicts from yesterday. Go into the challenges directory and try your hand at 03_analysis/A_json.py.

1.3   Authentication

Twitter controls access to their servers via a process of authentication and authorization. Authentication is how you let Twitter know who you are, in a way that is very hard to fake. Authorization is how the account owner (which will usually be yourself unless you are writing a Twitter app) controls what you are allowed to do in Twitter using their account. In Twitter, different levels of authorization require different levels of authentication.

Because we want to be able to interact with everything, we'll need the highest level of authorization and the strictest level of authentication. In Twitter, this means that we need two sets of ID's (called keys or tokens) and passwords (called secrets):

- consumer_key

- consumer_secret
- access_token_key
- access_token_secret

We'll provide some for you to use, but if you want to get your own you need to create an account on Twitter with a verified phone number. Then, while signed in to your Twitter account, go to: https://apps.twitter.com/. Follow the prompts to generate your keys and access tokens. Note that getting the second ID/password pair requires that you manually set the authorization level of your app.

We've stored our credentials in a separate file, which is smart. However, we have uploaded it to Github so that you have them too, which is not smart.

You should NEVER NEVER NEVER do this in real life.

We've stored it in YAML format, because it is more human-readible than JSON is. However, once it's inside Python, these data structures behave the same way.

```
In [3]: import yaml

        with open('../etc/creds.yml', 'r') as f:
            creds = yaml.load(f)
```

We're going to load these credentials into a requests module specifically designed for handling the flavor of authentication management that Twitter uses.

```
In [4]: from requests_oauthlib import OAuth1Session

        twitter = OAuth1Session(**creds)
```

That ** syntax we just used is called a "double splat" and is a python convenience function for converting the key-value pairs of a dictionary into keyword-argument pairs to pass to a function.

1.4 Accessing the API

Access to Twitter's API is organized through URLs called "endpoints". An endpoint is the location at which you can submit a request for Twitter to do something for you.

For example, the "endpoint" to search for specific kinds of tweets is at:

`https://api.twitter.com/1.1/search/tweets.json`

whereas posting new tweets is at:

`https://api.twitter.com/1.1/statuses/update.json`

For more information on the REST APIs, end points, and terms, check out: https://dev.twitter.com/rest/public. For the Streaming APIs: https://dev.twitter.com/streaming/overview.

All APIs on Twitter are "rate-limited" - this means that you are only allowed to ask a set number of questions per unit time (to keep their servers from being overloaded). This rate varies by endpoint and authorization, so be sure to check their developer site for the action you are trying to take.

For example, at the lowest level of authorization (Twitter calls this `application only`), you are allowed to make 450 search requests per 15 minute window, or about one every two seconds. At the highest level of authorization (Twitter calls this `user`) you can submit 180 requests every 15 minutes, or only about once every five seconds.

> side note - Google search is the worst rate-limiting I've ever seen, with an allowance of one hundred requests per day, or about once every *nine hundred seconds*

Let's try a couple of simple API queries. We're going to specify query parameters with `param`.

```
In [5]: search = "https://api.twitter.com/1.1/search/tweets.json"

        r = twitter.get(search, params={'q' : 'technology'})
```

This has returned an http response object, which contains data like whether or not the request succeeded:

```
In [6]: r.ok
```

```
Out[6]: True
```

You can also get the http response code, and the reason why Twitter sent you that code (these are all super important for controlling the flow of your program).

```
In [7]: r.status_code, r.reason
```

```
Out[7]: (200, 'OK')
```

The data that we asked Twitter to send us in r.content

```
In [8]: r.content
```

ouici t-50610s0u150uidsiteomsil('"s789094045e720822256610853'4'shandatesall112511114070044305612015673,6080078061004'049ar:aonlt1120560074053330

But that's not helpful. We can extract it in python's representation of json with the `json` method:

```
In [9]: r.json()
```

```
Out[9]: {'search_metadata': {'completed_in': 0.034,
          'count': 15,
          'max_id': 730515610493325312,
          'max_id_str': '730515610493325312',
          'next_results': '?max_id=730515587617558527&q=technology&include_entities=1',
          'query': 'technology',
          'refresh_url': '?since_id=730515610493325312&q=technology&include_entities=1',
          'since_id': 0,
          'since_id_str': '0'},
         'statuses': [{'contributors': None,
           'coordinates': None,
           'created_at': 'Wed May 11 21:51:08 +0000 2016',
           'entities': {'hashtags': [],
            'symbols': [],
            'urls': [{'display_url': 'dlvr.it/LHW909',
               'expanded_url': 'http://dlvr.it/LHW909',
               'indices': [78, 101],
               'url': 'https://t.co/sIp4aGfyLL'}],
            'user_mentions': []},
           'favorite_count': 0,
           'favorited': False,
           'geo': None,
           'id': 730515610493325312,
           'id_str': '730515610493325312',
           'in_reply_to_screen_name': None,
           'in_reply_to_status_id': None,
           'in_reply_to_status_id_str': None,
           'in_reply_to_user_id': None,
           'in_reply_to_user_id_str': None,
           'is_quote_status': False,
```

```
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': '<a href="http://dlvr.it" rel="nofollow">dlvr.it</a>',
'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:,
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Sat Nov 14 06:30:26 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Technology News from all over the web, clustered into one place.',
 'entities': {'description': {'urls': []}},
 'favourites_count': 0,
 'follow_request_sent': False,
 'followers_count': 2508,
 'following': False,
 'friends_count': 1187,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 89886458,
 'id_str': '89886458',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 185,
 'location': '',
 'name': 'Technology News',
 'notifications': False,
 'profile_background_color': 'EBEBEB',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/53713965/253_
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/537139
 'profile_background_tile': False,
 'profile_image_url': 'http://pbs.twimg.com/profile_images/719199830/globe_normal.png',
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/719199830/globe_normal.png',
 'profile_link_color': '0026BF',
 'profile_sidebar_border_color': 'DFDFDF',
 'profile_sidebar_fill_color': 'F3F3F3',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'TechCluster',
 'statuses_count': 640962,
 'time_zone': 'Eastern Time (US & Canada)',
 'url': None,
 'utc_offset': -14400,
 'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:08 +0000 2016',
 'entities': {'hashtags': [],
  'symbols': [],
```

    'urls': [{'display_url': 'dlvr.it/LHW91c',
      'expanded_url': 'http://dlvr.it/LHW91c',
      'indices': [78, 101],
      'url': 'https://t.co/qTB1PPvUeZ'}],
   'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515610493296642,
 'id_str': '730515610493296642',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'dlvr.it',
 'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:,
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Mon Jun 21 03:24:43 +0000 2010',
  'default_profile': False,
  'default_profile_image': False,
  'description': 'Designer, Twitter addict, Artist,Thinker,Writer,Engineer,party animal,gadget
  'entities': {'description': {'urls': []},
   'url': {'urls': [{'display_url': 'facebook.com/profile.php?id',
      'expanded_url': 'http://www.facebook.com/profile.php?id=100000419649950',
      'indices': [0, 22],
      'url': 'http://t.co/NhxIupnCUJ'}]}},
  'favourites_count': 19,
  'follow_request_sent': False,
  'followers_count': 6294,
  'following': False,
  'friends_count': 4620,
  'geo_enabled': False,
  'has_extended_profile': False,
  'id': 157865482,
  'id_str': '157865482',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': 'en',
  'listed_count': 70,
  'location': 'Mumbai, India',
  'name': 'BabinaSebastian',
  'notifications': False,
  'profile_background_color': '0099B9',
  'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme4/bg.gif',
  'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme4/bg.gif',

　　　　　　　'profile_background_tile': False,
　　　　　　　'profile_image_url': 'http://pbs.twimg.com/profile_images/1139122979/1_2_normal.jpg',
　　　　　　　'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1139122979/1_2_normal.jpg',
　　　　　　　'profile_link_color': '0099B9',
　　　　　　　'profile_sidebar_border_color': '5ED4DC',
　　　　　　　'profile_sidebar_fill_color': '95E8EC',
　　　　　　　'profile_text_color': '3C3940',
　　　　　　　'profile_use_background_image': True,
　　　　　　　'protected': False,
　　　　　　　'screen_name': 'SebPeter',
　　　　　　　'statuses_count': 263869,
　　　　　　　'time_zone': 'Hawaii',
　　　　　　　'url': 'http://t.co/NhxIupnCUJ',
　　　　　　　'utc_offset': -36000,
　　　　　　　'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:08 +0000 2016',
 'entities': {'hashtags': [],
 'symbols': [],
 'urls': [{'display_url': 'sco.lt/7tsaED',
   'expanded_url': 'http://sco.lt/7tsaED',
   'indices': [88, 111],
   'url': 'https://t.co/HKjVTZzp8U'}],
 'user_mentions': [{'id': 209484168,
   'id_str': '209484168',
   'indices': [79, 87],
   'name': 'Scoop.it',
   'screen_name': 'scoopit'}]},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515609243557888,
 'id_str': '730515609243557888',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'Scoop.it',
 'text': 'Educational Technology Guy: Technology I use on a Daily Basis - 2016 Version | @sco
 'truncated': False,
 'user': {'contributors_enabled': False,
 'created_at': 'Tue May 19 05:15:56 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Former teacher living in Portishead, UK. Managing Director of Cymru Care Tra

      'entities': {'description': {'urls': []},
       'url': {'urls': [{'display_url': 'edwardjgjones.co.uk',
          'expanded_url': 'http://www.edwardjgjones.co.uk',
          'indices': [0, 22],
          'url': 'http://t.co/z78xG1fUY2'}]}},
      'favourites_count': 16,
      'follow_request_sent': False,
      'followers_count': 212,
      'following': False,
      'friends_count': 125,
      'geo_enabled': False,
      'has_extended_profile': False,
      'id': 41060791,
      'id_str': '41060791',
      'is_translation_enabled': False,
      'is_translator': False,
      'lang': 'en',
      'listed_count': 67,
      'location': 'Portishead, UK',
      'name': 'Edward Jones',
      'notifications': False,
      'profile_background_color': 'C6E2EE',
      'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme2/bg.gif',
      'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme2/bg.gif',
      'profile_background_tile': False,
      'profile_banner_url': 'https://pbs.twimg.com/profile_banners/41060791/1348084813',
      'profile_image_url': 'http://pbs.twimg.com/profile_images/538364886976192512/oEYrj_2g_normal.
      'profile_image_url_https': 'https://pbs.twimg.com/profile_images/538364886976192512/oEYrj_2g_
      'profile_link_color': '1F98C7',
      'profile_sidebar_border_color': 'C6E2EE',
      'profile_sidebar_fill_color': 'DAECF4',
      'profile_text_color': '663B12',
      'profile_use_background_image': True,
      'protected': False,
      'screen_name': 'edwardjgjones',
      'statuses_count': 6386,
      'time_zone': 'London',
      'url': 'http://t.co/z78xG1fUY2',
      'utc_offset': 3600,
      'verified': False}},
    {'contributors': None,
     'coordinates': None,
     'created_at': 'Wed May 11 21:51:08 +0000 2016',
     'entities': {'hashtags': [{'indices': [0, 9], 'text': 'SuryaRay'},
       {'indices': [10, 16], 'text': 'Surya'},
       {'indices': [17, 26], 'text': 'SuryaRay'},
       {'indices': [27, 33], 'text': 'Surya'}],
      'symbols': [],
      'urls': [{'display_url': 'dlvr.it/LHW90G',
        'expanded_url': 'http://dlvr.it/LHW90G',
        'indices': [82, 105],
        'url': 'https://t.co/1OLI8uqHQ8'}],
      'user_mentions': [{'id': 181432391,
        'id_str': '181432391',

    'indices': [106, 115],
     'name': 'Surya Ray',
     'screen_name': 'SuryaRay'},
   {'id': 181432391,
     'id_str': '181432391',
     'indices': [116, 125],
     'name': 'Surya Ray',
     'screen_name': 'SuryaRay'},
   {'id': 853727768,
     'id_str': '853727768',
     'indices': [126, 136],
     'name': 'Surya Ray',
     'screen_name': 'SuryaRay3'}]},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515609226641408,
 'id_str': '730515609226641408',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'dlvr.it',
 'text': '#SuryaRay #Surya #SuryaRay #Surya Windows 10s Wi-Fi credential sharing is going htt
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Sat Sep 29 22:50:09 +0000 2012',
  'default_profile': True,
  'default_profile_image': False,
  'description': 'Quotes',
  'entities': {'description': {'urls': []}},
  'favourites_count': 0,
  'follow_request_sent': False,
  'followers_count': 1010,
  'following': False,
  'friends_count': 941,
  'geo_enabled': False,
  'has_extended_profile': False,
  'id': 853727768,
  'id_str': '853727768',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': 'en',
  'listed_count': 115,
  'location': '',
  'name': 'Surya Ray',

    'notifications': False,
    'profile_background_color': 'C0DEED',
    'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
    'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
    'profile_background_tile': False,
    'profile_image_url': 'http://pbs.twimg.com/profile_images/2663632924/433110b50cd06971545c3b6
    'profile_image_url_https': 'https://pbs.twimg.com/profile_images/2663632924/433110b50cd06971
    'profile_link_color': '0084B4',
    'profile_sidebar_border_color': 'C0DEED',
    'profile_sidebar_fill_color': 'DDEEF6',
    'profile_text_color': '333333',
    'profile_use_background_image': True,
    'protected': False,
    'screen_name': 'SuryaRay3',
    'statuses_count': 773290,
    'time_zone': None,
    'url': None,
    'utc_offset': None,
    'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:07 +0000 2016',
 'entities': {'hashtags': [],
  'media': [{'display_url': 'pic.twitter.com/YuTKvRJXez',
    'expanded_url': 'http://twitter.com/ba_06121988/status/730515608157069312/photo/1',
    'id': 730515607913824256,
    'id_str': '730515607913824256',
    'indices': [102, 125],
    'media_url': 'http://pbs.twimg.com/media/CiNQEzCVAAADMUZ.jpg',
    'media_url_https': 'https://pbs.twimg.com/media/CiNQEzCVAAADMUZ.jpg',
    'sizes': {'large': {'h': 505, 'resize': 'fit', 'w': 640},
     'medium': {'h': 473, 'resize': 'fit', 'w': 600},
     'small': {'h': 268, 'resize': 'fit', 'w': 340},
     'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
    'type': 'photo',
    'url': 'https://t.co/YuTKvRJXez'}],
  'symbols': [],
  'urls': [{'display_url': 'dlvr.it/LHW8zd',
    'expanded_url': 'http://dlvr.it/LHW8zd',
    'indices': [78, 101],
    'url': 'https://t.co/0k7evnNdWy'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515608157069312,
 'id_str': '730515608157069312',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',

'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'dlvr.it',
'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:,
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu Oct 08 15:57:08 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Very fun Loving guy. Like to make new friend nd hanging out. Ultimate Party
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'yahoo.com',
     'expanded_url': 'http://yahoo.com',
     'indices': [0, 22],
     'url': 'http://t.co/w6bWe5Hps0'}]}},
 'favourites_count': 1,
 'follow_request_sent': False,
 'followers_count': 59,
 'following': False,
 'friends_count': 479,
 'geo_enabled': False,
 'has_extended_profile': True,
 'id': 80878564,
 'id_str': '80878564',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 4,
 'location': 'Ahmedabad, Gujarat',
 'name': 'Benjamin Abraham',
 'notifications': False,
 'profile_background_color': '642D8B',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme10/bg.gif',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme10/bg.gif',
 'profile_background_tile': True,
 'profile_image_url': 'http://pbs.twimg.com/profile_images/2441986492/Me_207_normal.jpg',
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/2441986492/Me_207_normal.jpg
 'profile_link_color': '7870A1',
 'profile_sidebar_border_color': '65B0DA',
 'profile_sidebar_fill_color': '7AC3EE',
 'profile_text_color': '3D1957',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'ba_06121988',
 'statuses_count': 4091,
 'time_zone': None,
 'url': 'http://t.co/w6bWe5Hps0',
 'utc_offset': None,
 'verified': False}},
{'contributors': None,
 'coordinates': None,

'created_at': 'Wed May 11 21:51:07 +0000 2016',
'entities': {'hashtags': [],
 'symbols': [],
 'urls': [{'display_url': 'sco.lt/9Cdjmb',
   'expanded_url': 'http://sco.lt/9Cdjmb',
   'indices': [88, 111],
   'url': 'https://t.co/gFShN0vKv2'}],
 'user_mentions': [{'id': 209484168,
   'id_str': '209484168',
   'indices': [79, 87],
   'name': 'Scoop.it',
   'screen_name': 'scoopit'}]},
'favorite_count': 0,
'favorited': False,
'geo': None,
'id': 730515607276421120,
'id_str': '730515607276421120',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'Scoop.it',
'text': 'Educational Technology Guy: Technology I use on a Daily Basis - 2016 Version | @sco
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Tue Mar 02 06:19:05 +0000 2010',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'WebTeachers specialises in online educational resources, providing informat
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'webteachersuk.com',
     'expanded_url': 'http://www.webteachersuk.com',
     'indices': [0, 23],
     'url': 'https://t.co/2MVYVn3183'}]}},
 'favourites_count': 18,
 'follow_request_sent': False,
 'followers_count': 859,
 'following': False,
 'friends_count': 92,
 'geo_enabled': True,
 'has_extended_profile': False,
 'id': 118954344,
 'id_str': '118954344',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',

'listed_count': 134,
 'location': 'Portishead, UK',
 'name': 'WebTeachers',
 'notifications': False,
 'profile_background_color': '000000',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme13/bg.gif',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme13/bg.gif',
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/118954344/1453273914',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/1437239143/webteachers_new_logo_nor
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1437239143/webteachers_new_l
 'profile_link_color': '3B94D9',
 'profile_sidebar_border_color': '000000',
 'profile_sidebar_fill_color': '000000',
 'profile_text_color': '000000',
 'profile_use_background_image': False,
 'protected': False,
 'screen_name': 'WebTeachersUK',
 'statuses_count': 15243,
 'time_zone': 'London',
 'url': 'https://t.co/2MVYVn3183',
 'utc_offset': 3600,
 'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:07 +0000 2016',
 'entities': {'hashtags': [{'indices': [102, 107], 'text': 'IEEE'},
   {'indices': [108, 118], 'text': 'IEEEGreen'}],
  'symbols': [],
  'urls': [{'display_url': 'dlvr.it/LHW90C',
    'expanded_url': 'http://dlvr.it/LHW90C',
    'indices': [78, 101],
    'url': 'https://t.co/QuqB0ZJqEA'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515607171407872,
 'id_str': '730515607171407872',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'dlvr.it',
 'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:/
 'truncated': False,

      'user': {'contributors_enabled': False,
       'created_at': 'Thu Sep 16 10:43:45 +0000 2010',
       'default_profile': False,
       'default_profile_image': False,
       'description': 'University of Greenwich IEEE student branch (IEEEGreen)',
       'entities': {'description': {'urls': []},
        'url': {'urls': [{'display_url': 'ieeegreen.org',
           'expanded_url': 'http://www.ieeegreen.org/',
           'indices': [0, 22],
           'url': 'http://t.co/wKJyLLWcFD'}]}},
       'favourites_count': 0,
       'follow_request_sent': False,
       'followers_count': 757,
       'following': False,
       'friends_count': 6,
       'geo_enabled': True,
       'has_extended_profile': False,
       'id': 191396911,
       'id_str': '191396911',
       'is_translation_enabled': False,
       'is_translator': False,
       'lang': 'en',
       'listed_count': 137,
       'location': '',
       'name': 'ieeegreen',
       'notifications': False,
       'profile_background_color': '352726',
       'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/149629980/BG
       'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/149629
       'profile_background_tile': True,
       'profile_image_url': 'http://pbs.twimg.com/profile_images/1138510679/90_normal.JPG',
       'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1138510679/90_normal.JPG',
       'profile_link_color': 'D02B55',
       'profile_sidebar_border_color': '829D5E',
       'profile_sidebar_fill_color': '99CC33',
       'profile_text_color': '3E4415',
       'profile_use_background_image': True,
       'protected': False,
       'screen_name': 'IEEEGreen',
       'statuses_count': 343675,
       'time_zone': 'London',
       'url': 'http://t.co/wKJyLLWcFD',
       'utc_offset': 3600,
       'verified': False}},
     {'contributors': None,
      'coordinates': None,
      'created_at': 'Wed May 11 21:51:07 +0000 2016',
      'entities': {'hashtags': [{'indices': [25, 28], 'text': 'VR'},
        {'indices': [90, 93], 'text': 'VR'},
        {'indices': [94, 100], 'text': 'Gamer'},
        {'indices': [101, 106], 'text': 'Tech'},
        {'indices': [107, 112], 'text': 'News'},
        {'indices': [113, 127], 'text': 'QuaestorMovie'}],
       'media': [{'display_url': 'pic.twitter.com/qqibznEEtJ',

      'expanded_url': 'http://twitter.com/771h5/status/717382047921254400/photo/1',
      'id': 717382046620983297,
      'id_str': '717382046620983297',
      'indices': [139, 140],
      'media_url': 'http://pbs.twimg.com/media/CfSnK6kWIAEm-fi.jpg',
      'media_url_https': 'https://pbs.twimg.com/media/CfSnK6kWIAEm-fi.jpg',
      'sizes': {'large': {'h': 576, 'resize': 'fit', 'w': 1024},
       'medium': {'h': 338, 'resize': 'fit', 'w': 600},
       'small': {'h': 191, 'resize': 'fit', 'w': 340},
       'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
      'source_status_id': 717382047921254400,
      'source_status_id_str': '717382047921254400',
      'source_user_id': 177925681,
      'source_user_id_str': '177925681',
      'type': 'photo',
      'url': 'https://t.co/qqibznEEtJ'}],
    'symbols': [],
    'urls': [{'display_url': 'venturebeat.com/2016/04/05/the',
      'expanded_url': 'http://venturebeat.com/2016/04/05/the-htc-vive-offers-the-best-consumer-v
      'indices': [66, 89],
      'url': 'https://t.co/NdIM7o3wWl'}],
    'user_mentions': [{'id': 177925681,
      'id_str': '177925681',
      'indices': [3, 9],
      'name': 'Tweet Deck',
      'screen_name': '771h5'}]},
   'favorite_count': 0,
   'favorited': False,
   'geo': None,
   'id': 730515605443510272,
   'id_str': '730515605443510272',
   'in_reply_to_screen_name': None,
   'in_reply_to_status_id': None,
   'in_reply_to_status_id_str': None,
   'in_reply_to_user_id': None,
   'in_reply_to_user_id_str': None,
   'is_quote_status': False,
   'lang': 'en',
   'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
   'place': None,
   'possibly_sensitive': False,
   'retweet_count': 13,
   'retweeted': False,
   'retweeted_status': {'contributors': None,
    'coordinates': None,
    'created_at': 'Tue Apr 05 16:03:03 +0000 2016',
    'entities': {'hashtags': [{'indices': [14, 17], 'text': 'VR'},
      {'indices': [79, 82], 'text': 'VR'},
      {'indices': [83, 89], 'text': 'Gamer'},
      {'indices': [90, 95], 'text': 'Tech'},
      {'indices': [96, 101], 'text': 'News'},
      {'indices': [102, 116], 'text': 'QuaestorMovie'}],
     'media': [{'display_url': 'pic.twitter.com/qqibznEEtJ',
       'expanded_url': 'http://twitter.com/771h5/status/717382047921254400/photo/1',

```
        'id': 717382046620983297,
        'id_str': '717382046620983297',
        'indices': [117, 140],
        'media_url': 'http://pbs.twimg.com/media/CfSnK6kWIAEm-fi.jpg',
        'media_url_https': 'https://pbs.twimg.com/media/CfSnK6kWIAEm-fi.jpg',
        'sizes': {'large': {'h': 576, 'resize': 'fit', 'w': 1024},
         'medium': {'h': 338, 'resize': 'fit', 'w': 600},
         'small': {'h': 191, 'resize': 'fit', 'w': 340},
         'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
        'type': 'photo',
        'url': 'https://t.co/qqibznEEtJ'}],
   'symbols': [],
   'urls': [{'display_url': 'venturebeat.com/2016/04/05/the',
      'expanded_url': 'http://venturebeat.com/2016/04/05/the-htc-vive-offers-the-best-consumer-
      'indices': [55, 78],
      'url': 'https://t.co/NdIM7o3wWl'}],
   'user_mentions': []},
 'favorite_count': 18,
 'favorited': False,
 'geo': None,
 'id': 717382047921254400,
 'id_str': '717382047921254400',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 13,
 'retweeted': False,
 'source': '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
 'text': '$800 HTC best #VR experience - but needs better games  https://t.co/NdIM7o3wWl #VR
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Fri Aug 13 12:20:07 +0000 2010',
  'default_profile': False,
  'default_profile_image': False,
  'description': '#771h5 Film, video game and branded entertainment.',
  'entities': {'description': {'urls': []},
   'url': {'urls': [{'display_url': '771h5.com',
      'expanded_url': 'http://www.771h5.com',
      'indices': [0, 23],
      'url': 'https://t.co/JWsiK6NpfM'}]}},
  'favourites_count': 38675,
  'follow_request_sent': False,
  'followers_count': 50619,
  'following': False,
  'friends_count': 6999,
  'geo_enabled': False,
  'has_extended_profile': False,
```

  'id': 177925681,
  'id_str': '177925681',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': 'en',
  'listed_count': 953,
  'location': 'London, England. ',
  'name': 'Tweet Deck',
  'notifications': False,
  'profile_background_color': '000000',
  'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/564133965326
  'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/56413
  'profile_background_tile': False,
  'profile_banner_url': 'https://pbs.twimg.com/profile_banners/177925681/1432509344',
  'profile_image_url': 'http://pbs.twimg.com/profile_images/602614058114805761/PhIIHm6p_normal
  'profile_image_url_https': 'https://pbs.twimg.com/profile_images/602614058114805761/PhIIHm6p
  'profile_link_color': '000000',
  'profile_sidebar_border_color': '00060A',
  'profile_sidebar_fill_color': '01080A',
  'profile_text_color': 'E0D7E0',
  'profile_use_background_image': True,
  'protected': False,
  'screen_name': '771h5',
  'statuses_count': 2612,
  'time_zone': 'Madrid',
  'url': 'https://t.co/JWsiK6NpfM',
  'utc_offset': 7200,
  'verified': False}},
'source': 'Twitter Web Client',
'text': 'RT @771h5: $800 HTC best #VR experience - but needs better games https://t.co/NdIM
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Mon Oct 19 12:58:52 +0000 2015',
 'default_profile': False,
 'default_profile_image': False,
 'description': '2250. Civilisation is entering the third era. #SciFi #VR #MindFactories #77
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': '771h5.com',
     'expanded_url': 'http://www.771h5.com',
     'indices': [0, 23],
     'url': 'https://t.co/JWsiK6NpfM'}]}},
 'favourites_count': 929,
 'follow_request_sent': False,
 'followers_count': 15556,
 'following': False,
 'friends_count': 1636,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 3994538241,
 'id_str': '3994538241',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 179,

   'location': '',
   'name': 'Quaestor',
   'notifications': False,
   'profile_background_color': '000000',
   'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
   'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
   'profile_background_tile': False,
   'profile_banner_url': 'https://pbs.twimg.com/profile_banners/3994538241/1458868423',
   'profile_image_url': 'http://pbs.twimg.com/profile_images/666201681428918273/RfTcZUw1_normal
   'profile_image_url_https': 'https://pbs.twimg.com/profile_images/666201681428918273/RfTcZUw1_
   'profile_link_color': '0000FF',
   'profile_sidebar_border_color': '000000',
   'profile_sidebar_fill_color': '000000',
   'profile_text_color': '000000',
   'profile_use_background_image': False,
   'protected': False,
   'screen_name': 'QuaestorMovie',
   'statuses_count': 297,
   'time_zone': None,
   'url': 'https://t.co/JWsiK6NpfM',
   'utc_offset': None,
   'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Wed May 11 21:51:06 +0000 2016',
  'entities': {'hashtags': [{'indices': [102, 117],
     'text': 'TEAMFOLLOWBACK'}],
   'symbols': [],
   'urls': [{'display_url': 'dlvr.it/LHW8yg',
     'expanded_url': 'http://dlvr.it/LHW8yg',
     'indices': [78, 101],
     'url': 'https://t.co/5hyYailmnr'}],
   'user_mentions': []},
  'favorite_count': 1,
  'favorited': False,
  'geo': None,
  'id': 730515602771566593,
  'id_str': '730515602771566593',
  'in_reply_to_screen_name': None,
  'in_reply_to_status_id': None,
  'in_reply_to_status_id_str': None,
  'in_reply_to_user_id': None,
  'in_reply_to_user_id_str': None,
  'is_quote_status': False,
  'lang': 'en',
  'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
  'place': None,
  'possibly_sensitive': False,
  'retweet_count': 0,
  'retweeted': False,
  'source': 'dlvr.it',
  'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:,
  'truncated': False,
  'user': {'contributors_enabled': False,

        'created_at': 'Thu Aug 04 10:52:13 +0000 2011',
        'default_profile': False,
        'default_profile_image': False,
        'description': '#TEAMFOLLOWBACK',
        'entities': {'description': {'urls': []}},
        'favourites_count': 0,
        'follow_request_sent': False,
        'followers_count': 362,
        'following': False,
        'friends_count': 41,
        'geo_enabled': False,
        'has_extended_profile': False,
        'id': 348396292,
        'id_str': '348396292',
        'is_translation_enabled': False,
        'is_translator': False,
        'lang': 'ru',
        'listed_count': 53,
        'location': 'Canberra',
        'name': 'Happy FollowBack',
        'notifications': False,
        'profile_background_color': 'C0DEED',
        'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/304565518/fre
        'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/304565
        'profile_background_tile': False,
        'profile_image_url': 'http://pbs.twimg.com/profile_images/1477411465/7_normal.jpg',
        'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1477411465/7_normal.jpg',
        'profile_link_color': '0084B4',
        'profile_sidebar_border_color': 'C0DEED',
        'profile_sidebar_fill_color': 'DDEEF6',
        'profile_text_color': '333333',
        'profile_use_background_image': True,
        'protected': False,
        'screen_name': 'HappyFollowBack',
        'statuses_count': 307759,
        'time_zone': None,
        'url': None,
        'utc_offset': None,
        'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:06 +0000 2016',
 'entities': {'hashtags': [{'indices': [98, 103], 'text': 'Tech'},
   {'indices': [104, 110], 'text': 'Apple'},
   {'indices': [111, 115], 'text': 'Mac'}],
  'symbols': [],
  'urls': [{'display_url': 'goo.gl/fb/0j4CEe',
    'expanded_url': 'http://goo.gl/fb/0j4CEe',
    'indices': [74, 97],
    'url': 'https://t.co/CbVMdnN1ll'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,

'id': 730515600167022596,
'id_str': '730515600167022596',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'Google',
'text': 'New Siri sibling Viv may be next step in AI evolution: With the creators https://t.
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Fri Mar 16 05:59:50 +0000 2012',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Graphic / Web Design',
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'jedigraphix.com',
     'expanded_url': 'http://www.jedigraphix.com',
     'indices': [0, 22],
     'url': 'http://t.co/7md9pI7LUR'}]}},
 'favourites_count': 0,
 'follow_request_sent': False,
 'followers_count': 94,
 'following': False,
 'friends_count': 53,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 526081857,
 'id_str': '526081857',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 152,
 'location': '',
 'name': 'Jedi Graphix',
 'notifications': False,
 'profile_background_color': 'FFFFFF',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/454582632/jed
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/454582
 'profile_background_tile': True,
 'profile_image_url': 'http://pbs.twimg.com/profile_images/1914283209/jedilogo_normal.jpg',
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1914283209/jedilogo_normal.j
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'C0DEED',
 'profile_sidebar_fill_color': 'DDEEF6',
 'profile_text_color': '333333',
 'profile_use_background_image': True,

        'protected': False,
        'screen_name': 'jedigraphix',
        'statuses_count': 217178,
        'time_zone': None,
        'url': 'http://t.co/7md9pI7LUR',
        'utc_offset': None,
        'verified': False}},
    {'contributors': None,
     'coordinates': None,
     'created_at': 'Wed May 11 21:51:04 +0000 2016',
     'entities': {'hashtags': [],
      'symbols': [],
      'urls': [{'display_url': 'nytimes.com/2016/05/06/tec',
          'expanded_url': 'http://www.nytimes.com/2016/05/06/technology/personaltech/fixing-a-login-
          'indices': [58, 81],
          'url': 'https://t.co/xysXyzyCXu'}],
      'user_mentions': []},
     'favorite_count': 0,
     'favorited': False,
     'geo': None,
     'id': 730515594995499009,
     'id_str': '730515594995499009',
     'in_reply_to_screen_name': None,
     'in_reply_to_status_id': None,
     'in_reply_to_status_id_str': None,
     'in_reply_to_user_id': None,
     'in_reply_to_user_id_str': None,
     'is_quote_status': False,
     'lang': 'en',
     'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
     'place': None,
     'possibly_sensitive': False,
     'retweet_count': 0,
     'retweeted': False,
     'source': 'IFTTT',
     'text': 'Tech Tip: Fixing a Login Keychain Password Error in OSX https://t.co/xysXyzyCXu',
     'truncated': False,
     'user': {'contributors_enabled': False,
      'created_at': 'Mon Jan 11 06:35:25 +0000 2010',
      'default_profile': False,
      'default_profile_image': False,
      'description': 'Digital Marketer by the day, Tech geek by the night!\na generous movie buf
      'entities': {'description': {'urls': []}},
      'favourites_count': 99,
      'follow_request_sent': False,
      'followers_count': 500,
      'following': False,
      'friends_count': 1205,
      'geo_enabled': True,
      'has_extended_profile': False,
      'id': 103779751,
      'id_str': '103779751',
      'is_translation_enabled': False,
      'is_translator': False,

      'lang': 'en',
      'listed_count': 48,
      'location': 'New Delhi',
      'name': 'ashish sahani',
      'notifications': False,
      'profile_background_color': '0F0A02',
      'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/5407586919856
      'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/540758
      'profile_background_tile': False,
      'profile_banner_url': 'https://pbs.twimg.com/profile_banners/103779751/1451628144',
      'profile_image_url': 'http://pbs.twimg.com/profile_images/539480965744050177/d6PIzBIs_normal
      'profile_image_url_https': 'https://pbs.twimg.com/profile_images/539480965744050177/d6PIzBIs_
      'profile_link_color': '000000',
      'profile_sidebar_border_color': 'BCB302',
      'profile_sidebar_fill_color': '171106',
      'profile_text_color': '8A7302',
      'profile_use_background_image': True,
      'protected': False,
      'screen_name': 'digihani',
      'statuses_count': 15863,
      'time_zone': 'Samoa',
      'url': None,
      'utc_offset': -39600,
      'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Wed May 11 21:51:04 +0000 2016',
  'entities': {'hashtags': [{'indices': [58, 63], 'text': 'tech'}],
   'symbols': [],
   'urls': [{'display_url': 'ift.tt/1rworqU',
     'expanded_url': 'http://ift.tt/1rworqU',
     'indices': [64, 87],
     'url': 'https://t.co/cFaTlVBcNi'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,
  'geo': None,
  'id': 730515594341171200,
  'id_str': '730515594341171200',
  'in_reply_to_screen_name': None,
  'in_reply_to_status_id': None,
  'in_reply_to_status_id_str': None,
  'in_reply_to_user_id': None,
  'in_reply_to_user_id_str': None,
  'is_quote_status': False,
  'lang': 'en',
  'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
  'place': None,
  'possibly_sensitive': False,
  'retweet_count': 0,
  'retweeted': False,
  'source': 'IFTTT',
  'text': 'Tech Tip: Fixing a Login Keychain Password Error in OSX #tech https://t.co/cFaTlVBcl
  'truncated': False,

    'user': {'contributors_enabled': False,
     'created_at': 'Thu Nov 05 21:52:06 +0000 2015',
     'default_profile': True,
     'default_profile_image': False,
     'description': '',
     'entities': {'description': {'urls': []}},
     'favourites_count': 5,
     'follow_request_sent': False,
     'followers_count': 18,
     'following': False,
     'friends_count': 41,
     'geo_enabled': False,
     'has_extended_profile': False,
     'id': 4122311729,
     'id_str': '4122311729',
     'is_translation_enabled': False,
     'is_translator': False,
     'lang': 'en',
     'listed_count': 31,
     'location': '',
     'name': 'Mark Troy',
     'notifications': False,
     'profile_background_color': 'C0DEED',
     'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
     'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
     'profile_background_tile': False,
     'profile_banner_url': 'https://pbs.twimg.com/profile_banners/4122311729/1446761348',
     'profile_image_url': 'http://pbs.twimg.com/profile_images/662388486540910592/ZcOH27eq_normal
     'profile_image_url_https': 'https://pbs.twimg.com/profile_images/662388486540910592/ZcOH27eq_
     'profile_link_color': '0084B4',
     'profile_sidebar_border_color': 'C0DEED',
     'profile_sidebar_fill_color': 'DDEEF6',
     'profile_text_color': '333333',
     'profile_use_background_image': True,
     'protected': False,
     'screen_name': 'MarkTroy9',
     'statuses_count': 1444,
     'time_zone': None,
     'url': None,
     'utc_offset': None,
     'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Wed May 11 21:51:03 +0000 2016',
  'entities': {'hashtags': [],
   'media': [{'display_url': 'pic.twitter.com/M4QBbnTHZ9',
     'expanded_url': 'http://twitter.com/jonasolvr/status/730464563808833536/photo/1',
     'id': 730464560768028672,
     'id_str': '730464560768028672',
     'indices': [48, 71],
     'media_url': 'http://pbs.twimg.com/media/CiMhpdmXEAAFcj7.jpg',
     'media_url_https': 'https://pbs.twimg.com/media/CiMhpdmXEAAFcj7.jpg',
     'sizes': {'large': {'h': 408, 'resize': 'fit', 'w': 289},
      'medium': {'h': 408, 'resize': 'fit', 'w': 289},

          'small': {'h': 408, 'resize': 'fit', 'w': 289},
          'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
       'source_status_id': 730464563808833536,
       'source_status_id_str': '730464563808833536',
       'source_user_id': 207697493,
       'source_user_id_str': '207697493',
       'type': 'photo',
       'url': 'https://t.co/M4QBbnTHZ9'}],
     'symbols': [],
     'urls': [],
     'user_mentions': [{'id': 207697493,
       'id_str': '207697493',
       'indices': [3, 13],
       'name': 'jonas',
       'screen_name': 'jonasolvr'}]},
   'favorite_count': 0,
   'favorited': False,
   'geo': None,
   'id': 730515591501582336,
   'id_str': '730515591501582336',
   'in_reply_to_screen_name': None,
   'in_reply_to_status_id': None,
   'in_reply_to_status_id_str': None,
   'in_reply_to_user_id': None,
   'in_reply_to_user_id_str': None,
   'is_quote_status': False,
   'lang': 'und',
   'metadata': {'iso_language_code': 'und', 'result_type': 'recent'},
   'place': None,
   'possibly_sensitive': False,
   'retweet_count': 10,
   'retweeted': False,
   'retweeted_status': {'contributors': None,
     'coordinates': None,
     'created_at': 'Wed May 11 18:28:18 +0000 2016',
     'entities': {'hashtags': [],
       'media': [{'display_url': 'pic.twitter.com/M4QBbnTHZ9',
         'expanded_url': 'http://twitter.com/jonasolvr/status/730464563808833536/photo/1',
         'id': 730464560768028672,
         'id_str': '730464560768028672',
         'indices': [33, 56],
         'media_url': 'http://pbs.twimg.com/media/CiMhpdmXEAAFcj7.jpg',
         'media_url_https': 'https://pbs.twimg.com/media/CiMhpdmXEAAFcj7.jpg',
         'sizes': {'large': {'h': 408, 'resize': 'fit', 'w': 289},
          'medium': {'h': 408, 'resize': 'fit', 'w': 289},
          'small': {'h': 408, 'resize': 'fit', 'w': 289},
          'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
         'type': 'photo',
         'url': 'https://t.co/M4QBbnTHZ9'}],
       'symbols': [],
       'urls': [],
       'user_mentions': []},
     'favorite_count': 5,
     'favorited': False,

'geo': None,
'id': 730464563808833536,
'id_str': '730464563808833536',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'und',
'metadata': {'iso_language_code': 'und', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 10,
'retweeted': False,
'source': 'Twitter Web Client',
'text': 'tipo citroen creative technology https://t.co/M4QBbnTHZ9',
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Mon Oct 25 21:28:41 +0000 2010',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'snap: pfvrjonas',
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'instagram.com/jonasolvr',
     'expanded_url': 'http://instagram.com/jonasolvr',
     'indices': [0, 23],
     'url': 'https://t.co/n7hbcOxjDN'}]}},
 'favourites_count': 4778,
 'follow_request_sent': False,
 'followers_count': 3660,
 'following': False,
 'friends_count': 649,
 'geo_enabled': True,
 'has_extended_profile': True,
 'id': 207697493,
 'id_str': '207697493',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'pt',
 'listed_count': 8,
 'location': '',
 'name': 'jonas',
 'notifications': False,
 'profile_background_color': 'FFFFFF',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/643962924/15
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/64396
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/207697493/1461799239',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/729189620219584512/P4LBiTK-_normal
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/729189620219584512/P4LBiTK-
 'profile_link_color': '000000',
 'profile_sidebar_border_color': 'FFFFFF',
 'profile_sidebar_fill_color': 'EFEFEF',

    'profile_text_color': '333333',
    'profile_use_background_image': False,
    'protected': False,
    'screen_name': 'jonasolvr',
    'statuses_count': 84254,
    'time_zone': 'Brasilia',
    'url': 'https://t.co/n7hbcOxjDN',
    'utc_offset': -10800,
    'verified': False}},
 'source': 'Twitter for iPhone
 'text': 'RT @jonasolvr: tipo citroen creative technology https://t.co/M4QBbnTHZ9',
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Sat Oct 18 00:17:00 +0000 2014',
  'default_profile': False,
  'default_profile_image': False,
  'description': 'S fique preso ao que te liberta.',
  'entities': {'description': {'urls': []}},
  'favourites_count': 1681,
  'follow_request_sent': False,
  'followers_count': 270,
  'following': False,
  'friends_count': 341,
  'geo_enabled': True,
  'has_extended_profile': True,
  'id': 2861511404,
  'id_str': '2861511404',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': 'pt',
  'listed_count': 0,
  'location': 'Rio de Janeiro, Brasil',
  'name': 'Linda',
  'notifications': False,
  'profile_background_color': '000000',
  'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
  'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
  'profile_background_tile': False,
  'profile_banner_url': 'https://pbs.twimg.com/profile_banners/2861511404/1462076249',
  'profile_image_url': 'http://pbs.twimg.com/profile_images/726626570404769794/pgNaAy3n_normal
  'profile_image_url_https': 'https://pbs.twimg.com/profile_images/726626570404769794/pgNaAy3n_
  'profile_link_color': '3B94D9',
  'profile_sidebar_border_color': '000000',
  'profile_sidebar_fill_color': '000000',
  'profile_text_color': '000000',
  'profile_use_background_image': False,
  'protected': False,
  'screen_name': 'lindabackx_',
  'statuses_count': 3831,
  'time_zone': None,
  'url': None,
  'utc_offset': None,
  'verified': False}},
{'contributors': None,

'coordinates': None,
'created_at': 'Wed May 11 21:51:03 +0000 2016',
'entities': {'hashtags': [],
 'symbols': [],
 'urls': [{'display_url': 'ift.tt/21Y1wBk',
   'expanded_url': 'http://ift.tt/21Y1wBk',
   'indices': [78, 101],
   'url': 'https://t.co/UW0Vuuohto'}],
 'user_mentions': []},
'favorite_count': 0,
'favorited': False,
'geo': None,
'id': 730515588964093957,
'id_str': '730515588964093957',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'IFTTT',
'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:/
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Wed Mar 26 12:05:33 +0000 2008',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Tecnpata, estudioso de la tecnologa. CEO de Samurai, agencia que integra te
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'samurai.cl',
     'expanded_url': 'http://www.samurai.cl',
     'indices': [0, 23],
     'url': 'https://t.co/SunueTxn4k'}]}},
 'favourites_count': 61,
 'follow_request_sent': False,
 'followers_count': 10325,
 'following': False,
 'friends_count': 7762,
 'geo_enabled': True,
 'has_extended_profile': True,
 'id': 14225006,
 'id_str': '14225006',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 208,
 'location': 'Santiago, Chile',
 'name': 'Mario Saavedra',

    'notifications': False,
    'profile_background_color': '3B94D9',
    'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/3788000000298
    'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/378800
    'profile_background_tile': True,
    'profile_banner_url': 'https://pbs.twimg.com/profile_banners/14225006/1458713619',
    'profile_image_url': 'http://pbs.twimg.com/profile_images/501719544319070208/gMwulhYl_normal
    'profile_image_url_https': 'https://pbs.twimg.com/profile_images/501719544319070208/gMwulhYl_
    'profile_link_color': 'CC3366',
    'profile_sidebar_border_color': 'FFFFFF',
    'profile_sidebar_fill_color': 'EFEFEF',
    'profile_text_color': '333333',
    'profile_use_background_image': False,
    'protected': False,
    'screen_name': 'MacGenio',
    'statuses_count': 109872,
    'time_zone': 'Santiago',
    'url': 'https://t.co/SunueTxn4k',
    'utc_offset': -10800,
    'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Wed May 11 21:51:03 +0000 2016',
 'entities': {'hashtags': [{'indices': [50, 57], 'text': 'Python'}],
  'symbols': [],
  'urls': [{'display_url': 'j.mp/21Ef0Ck',
    'expanded_url': 'http://j.mp/21Ef0Ck',
    'indices': [103, 126],
    'url': 'https://t.co/GdbBaxk5Or'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 730515587617558528,
 'id_str': '730515587617558528',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': '<a href="http://techeventsnetwork.com/" rel="nofollow">Tech Events Network - NYC<,
 'text': 'Going to the Byte Academy: FinTech, Data Science, #Python Coding Bootcamp meetup ne
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Sun Feb 07 21:43:30 +0000 2016',
  'default_profile': True,
  'default_profile_image': False,

```
                'description': "Events for developers, technologists, and other geeks in New York City. If
                'entities': {'description': {'urls': []}},
                'favourites_count': 1,
                'follow_request_sent': False,
                'followers_count': 195,
                'following': False,
                'friends_count': 734,
                'geo_enabled': False,
                'has_extended_profile': False,
                'id': 4875487440,
                'id_str': '4875487440',
                'is_translation_enabled': False,
                'is_translator': False,
                'lang': 'en',
                'listed_count': 61,
                'location': 'New York, NY',
                'name': 'NYC Tech Events',
                'notifications': False,
                'profile_background_color': 'F5F8FA',
                'profile_background_image_url': None,
                'profile_background_image_url_https': None,
                'profile_background_tile': False,
                'profile_banner_url': 'https://pbs.twimg.com/profile_banners/4875487440/1455209318',
                'profile_image_url': 'http://pbs.twimg.com/profile_images/698639359512195072/nmORkXm9_normal
                'profile_image_url_https': 'https://pbs.twimg.com/profile_images/698639359512195072/nmORkXm9_
                'profile_link_color': '2B7BB9',
                'profile_sidebar_border_color': 'C0DEED',
                'profile_sidebar_fill_color': 'DDEEF6',
                'profile_text_color': '333333',
                'profile_use_background_image': True,
                'protected': False,
                'screen_name': 'NYC_TechEvents',
                'statuses_count': 1762,
                'time_zone': 'Eastern Time (US & Canada)',
                'url': None,
                'utc_offset': -14400,
                'verified': False}}]}
```

This has some helpful metadata about our request, like a url where we can get the next batch of results
from Twitter for the same query:

```
In [10]: data = r.json()
         data['search_metadata']

Out[10]: {'completed_in': 0.034,
          'count': 15,
          'max_id': 730515610493325312,
          'max_id_str': '730515610493325312',
          'next_results': '?max_id=730515587617558527&q=technology&include_entities=1',
          'query': 'technology',
          'refresh_url': '?since_id=730515610493325312&q=technology&include_entities=1',
          'since_id': 0,
          'since_id_str': '0'}
```

The tweets that we want are under the key "statuses"

```python
In [11]: statuses = data['statuses']
         statuses[0]
```

```
Out[11]: {'contributors': None,
          'coordinates': None,
          'created_at': 'Wed May 11 21:51:08 +0000 2016',
          'entities': {'hashtags': [],
           'symbols': [],
           'urls': [{'display_url': 'dlvr.it/LHW909',
             'expanded_url': 'http://dlvr.it/LHW909',
             'indices': [78, 101],
             'url': 'https://t.co/sIp4aGfyLL'}],
           'user_mentions': []},
          'favorite_count': 0,
          'favorited': False,
          'geo': None,
          'id': 730515610493325312,
          'id_str': '730515610493325312',
          'in_reply_to_screen_name': None,
          'in_reply_to_status_id': None,
          'in_reply_to_status_id_str': None,
          'in_reply_to_user_id': None,
          'in_reply_to_user_id_str': None,
          'is_quote_status': False,
          'lang': 'en',
          'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
          'place': None,
          'possibly_sensitive': False,
          'retweet_count': 0,
          'retweeted': False,
          'source': '<a href="http://dlvr.it" rel="nofollow">dlvr.it</a>',
          'text': 'Windows 10s Wi-Fi credential sharing is going away in the Anniversary Update https:/,
          'truncated': False,
          'user': {'contributors_enabled': False,
           'created_at': 'Sat Nov 14 06:30:26 +0000 2009',
           'default_profile': False,
           'default_profile_image': False,
           'description': 'Technology News from all over the web, clustered into one place.',
           'entities': {'description': {'urls': []}},
           'favourites_count': 0,
           'follow_request_sent': False,
           'followers_count': 2508,
           'following': False,
           'friends_count': 1187,
           'geo_enabled': False,
           'has_extended_profile': False,
           'id': 89886458,
           'id_str': '89886458',
           'is_translation_enabled': False,
           'is_translator': False,
           'lang': 'en',
           'listed_count': 185,
           'location': '',
           'name': 'Technology News',
```

```
'notifications': False,
'profile_background_color': 'EBEBEB',
'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/53713965/253_P
'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/5371396
'profile_background_tile': False,
'profile_image_url': 'http://pbs.twimg.com/profile_images/719199830/globe_normal.png',
'profile_image_url_https': 'https://pbs.twimg.com/profile_images/719199830/globe_normal.png',
'profile_link_color': '0026BF',
'profile_sidebar_border_color': 'DFDFDF',
'profile_sidebar_fill_color': 'F3F3F3',
'profile_text_color': '333333',
'profile_use_background_image': True,
'protected': False,
'screen_name': 'TechCluster',
'statuses_count': 640962,
'time_zone': 'Eastern Time (US & Canada)',
'url': None,
'utc_offset': -14400,
'verified': False}}
```

This is one tweet.

> Depending on which tweet this is, you may or may not see that Twitter automatically pulls out links and mentions and gives you their index location in the raw tweet string

Twitter gives you a whole lot of information about their users, including geographical coordinates, the device they are tweeting from, and links to their photographs.

Twitter supports what it calls query operators, which modify the search behavior. For example, if you want to search for tweets where a particular user is mentioned, include the at-sign, `@`, followed by the username. To search for tweets sent to a particular user, use `to:username`. For tweets from a particular user, `from:username`. For hashtags, use `#hashtag`.

For a complete set of options: https://dev.twitter.com/rest/public/search.

Let's try a more complicated search:

```
In [12]: r = twitter.get(search, params={
                'q' : 'happy',
                'geocode' : '37.8734855,-122.2597169,10mi'
            })
         r.ok

Out[12]: True

In [13]: statuses = r.json()['statuses']
         statuses[0]

Out[13]: {'contributors': None,
          'coordinates': None,
          'created_at': 'Wed May 11 21:51:07 +0000 2016',
          'entities': {'hashtags': [], 'symbols': [], 'urls': [], 'user_mentions': []},
          'favorite_count': 0,
          'favorited': False,
          'geo': None,
          'id': 730515605875396608,
          'id_str': '730515605875396608',
          'in_reply_to_screen_name': None,
          'in_reply_to_status_id': None,
```

'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'retweet_count': 0,
'retweeted': False,
'source': 'Twitter Web Client',
'text': 'That I get to give lectures combining cryptography and Sylvia Plath makes me very, v
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Sat Apr 17 01:47:46 +0000 2010',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Software dev & feminist educator. VP of Education at @Hackbright Academy. Py
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'joelburton.com',
     'expanded_url': 'http://joelburton.com',
     'indices': [0, 22],
     'url': 'http://t.co/sMlYeSB5Pf'}]}},
 'favourites_count': 5541,
 'follow_request_sent': False,
 'followers_count': 779,
 'following': False,
 'friends_count': 752,
 'geo_enabled': True,
 'has_extended_profile': False,
 'id': 133939475,
 'id_str': '133939475',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 50,
 'location': 'San Francisco',
 'name': 'Joel Burton',
 'notifications': False,
 'profile_background_color': '000000',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_tile': False,
 'profile_image_url': 'http://pbs.twimg.com/profile_images/1217917608/IMG_3419_normal.jpg',
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1217917608/IMG_3419_normal.jp
 'profile_link_color': 'DD2E44',
 'profile_sidebar_border_color': '000000',
 'profile_sidebar_fill_color': '000000',
 'profile_text_color': '000000',
 'profile_use_background_image': False,
 'protected': False,
 'screen_name': 'wjoelburton',
 'statuses_count': 1673,
 'time_zone': 'Pacific Time (US & Canada)',
 'url': 'http://t.co/sMlYeSB5Pf',

```
          'utc_offset': -25200,
          'verified': False}}
```

If we want to store this data somewhere, we can output it as json using the json library from above. However, if you're doing a lot of these, you'll probaby want to use a database to handle everything.

```
In [14]: with open('my_tweets.json', 'w') as f:
             json.dump(statuses, f)
```

To post tweets, we need to use a different endpoint:

```
In [15]: post = "https://api.twitter.com/1.1/statuses/update.json"
```

And now we can pass a new tweet (remember, Twitter calls these 'statuses') as a parameter to our post request.

```
In [16]: r = twitter.post(post, params={
                 'status' : "I stole Juan's Twitter credentials"
             })
         r.ok
```

```
Out[16]: False
```

Other (optional) parameters include things like location, and replies.

1.5 Scheduling

The real beauty of bots is that they are designed to work without interaction or oversight. Imagine a situation where you want to automatically retweet everything coming out of the D-Lab's twitter account, "@DLabAtBerkeley". You could:

1. spend the rest of your life glued to D-Lab's twitter page and hitting refresh; or,
2. write a function

We're going to import a module called `time` that will pause our code, so that we don't hit Twitter's rate limit

```
In [17]: import time

         def retweet():
             r = twitter.get(search, {'q':'DLabAtBerkeley'})
             if r.ok:
                 statuses = r.json()['statuses']
                 for update in statuses:
                     username = item['user']['screen_name']
                     parameters = {'status':'HOORAY! @' + username}
                     r = twitter.post(post, parameters)
                     print(r.status_code, r.reason)
                     time.sleep(5)
```

But you are a human that needs to eat, sleep, and be social with other humans. Luckily, Linux systems have a time-based daemon called `cron` that will run scripts like this *for you*.

People on windows and macs will not be able to run this. That's okay.

The way that `cron` works is it reads in files where each line has a time followed by a job (these are called cronjobs). You can edit your crontab by typing `crontab -e` into a terminal.

They looks like this:

```
In [18]: with open('../etc/crontab_example', 'r') as f:
             print(f.read())

# In a user's crontab, jobs run under that user
# Time is specified as <min> <hour> <day> <month> <wday>
# To specify any time, use '*'
# For unknown reasons, cronjobs fail unless the tab ends with a newline

00 08 * * 1 echo "It is 8am on Monday" >> /var/dumblog
```

This is telling `cron` to print that statement to a file called "dumblog" at 8am every Monday.

It's generally frowned upon to enter jobs through crontabs because they are hard to modify without breaking them. The better solution is to put your timed command into a file and copy the file into `/etc/cron.d/`. These files look like this:

```
In [19]: with open('../etc/crond_example', 'r') as f:
             print(f.read())

#!/bin/bash
# First, make sure you specify all of the paths that you might need to run
# your task. If you aren't sure, copy the entire $PATH variable

PATH=/home/dillon/.conda/envs/py27/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin

# Then, specify when you want the task to occur; the user account to run it;
# and the job

@hourly dillon cd ~/scripts; python simple.py
```

At this point, you might be a little upset that you can't do this on your laptop, but the truth is you don't really want to run daemons and cronjobs on your laptop, which goes to sleep and runs out of batteries. This is what servers are for (like AWS).

1.6 Now it is time for you to look at a Twitter bot

We've put a template in the `scripts` folder for a basic, responsive Twitter bot. If there's time, you can try it out, but be generous with your `time.sleep()` calls as the whole class is sharing this account. The user credentials come from @juanshishido, so please don't get him banned from Twitter.

If you have tried to run this, or some of the earlier code in this notebook, you have probably encountered some of Twitter's error codes. Here are the most common, and why you are triggering them.

1. `400 = bad request` - This means the API (middleman) doesn't like how you formatted your request. Check the API documentation to make sure you are doing things correctly.

2. `401 = unauthorized` - This either means you entered your auth codes incorrectly, or those auth codes don't have permission to do what you're trying to do. It takes Twitter a while to assign posting rights to your auth tokens after you've given them your phone number. If you have just done this, wait five minutes, then try again.

3. `403 = forbidden` - Twitter won't let you post what you are trying to post, most likely because you are trying to post the same tweet twice in a row within a few minutes of each other. Try changing your status update. If that doesn't fix it, then you are either:

   A. Hitting Twitter's daily posting limit. They don't say what this is.

   B. Trying to follow too many people, rapidly following and unfollowing the same person, or are otherwise making Twitter think you are a spambot

4. `429 = too many requests` - This means that you have exceeded Twitter's rate limit for whatever it is you are trying to do. Increase your `time.sleep()` value.

2 Tabular data

In data storage, data visualization, inferential statistics, and machine learning, the most common way to pass data between applications is in the form of tables (these are called tabular, structured, or rectangular data). These are convenient in that, when used correctly, they store data in a DRY and easily queryable way, and are also easily turned into matrices for numeric processing.

> note - it is sometimes tempting to refer to N-dimensional matrices as arrays, following the numpy naming convention, but these are not the same as arrays in C++ or Java, which may cause confusion

It is common in enterprise applications to store tabular data in a SQL database. In the sciences, data is typically passed around as comma separated value files (.csv), which you have already been dealing with over the course of the last two days.

For this brief introduction to analyzing tabular data, we'll be using the scipy stack, which includes numpy, pandas, scipy, and "scikits" like sk-learn and sk-image.

```
In [20]: import pandas as pd
```

You might not have seen this `as` convention yet. It is just telling python that when we import `pandas`, we don't want to access it in the namespace as `pandas` but as `pd` instead.

2.1 Pandas basics

We'll start by making a small table to practice on. Tables in pandas are called data frames, so we'll start by making an instance of class `DataFrame`, and initialize it with some data.

> note - pandas and R use the same name for their tables, but their behavior is often very different

```
In [21]: table = pd.DataFrame({'id': [1,2,3], 'name':['dillon','juan','andrew'], 'age':[47,27,23]})
         print(table)

age  id    name
0   47   1  dillon
1   27   2    juan
2   23   3  andrew
```

Variables in pandas are represented by a pandas-specific data structure, called a `Series`. You can grab a `Series` out of a `DataFrame` by using the slicing operator with the name of the variable that you want to pull.

```
In [22]: table['name'], type(table['name'])

Out[22]: (0     dillon
          1       juan
          2     andrew
          Name: name, dtype: object, pandas.core.series.Series)
```

We could have made each variable a `Series`, and then put it into the DataFrame object, but it's easier in this instance to pass in a dictionary where the keys are variable names and the values are lists. You can also modify a data frame in place using similar syntax:

```
In [23]: table['fingers'] = [9, 10, None]
```

If you try to run that code without the `None` there, pandas will return an error. In a table (in any language) each column must have the same number of rows.

We've entered `None`, base python's missingness indicator, but pandas is going to swap this out with something else:

```
In [24]: table['fingers']

Out[24]: 0      9
         1     10
         2    NaN
         Name: fingers, dtype: float64
```

You might be tempted to write your own control structures around these missing values (which are variably called NaN, nan, and NA), but this is always a bad idea:

```
In [25]: table['fingers'][2] == None

Out[25]: False

In [26]: table['fingers'][2] == 'NaN'

Out[26]: False

In [27]: type(table['fingers'][2]) == str

Out[27]: False
```

None of this works because the pandas NaN is a subclass of numpy's double precision floating point number. However, for ambiguous reasons, even numpy.nan does not evaluate as being equal to itself.

To handle missing data, you'll need to use the pandas method isnull.

```
In [28]: pd.isnull(table['fingers'])

Out[28]: 0    False
         1    False
         2     True
         Name: fingers, dtype: bool
```

In the same way that we've been pulling out columns by name, you can pull out rows by index. If I want to grab the first row, I can use:

```
In [29]: table[:1]

Out[29]:    age  id    name  fingers
         0   47   1  dillon        9
```

Recall that indices in python start at zero, and that selecting by a range does not include the final value (i.e. [,)).

Unlike other software languages (R, I'm looking at you here), row indices in pandas are immutable. So, if I rearrange my data, the index also get shuffled.

```
In [30]: table.sort_values('age')

Out[30]:    age  id    name  fingers
         2   23   3  andrew      NaN
         1   27   2    juan       10
         0   47   1  dillon        9
```

Because of this, it's common to set the index to be something like a timestamp or UUID.

We can select parts of a DataFrame with conditional statements:

```
In [31]: table[table['age'] < 40]

Out[31]:    age  id    name  fingers
         1   27   2    juan       10
         2   23   3  andrew      NaN
```

37

2.2 Merging tables

As you might expect, tables in pandas can also be merged by keys. So, if we make a new dataset that shares an attribute in common:

```
In [32]: other_table = pd.DataFrame({
                'name':['dav', 'juan', 'dillon'],
                'languages':['python','python','python']})

In [33]: table.merge(other_table, on='name')

Out[33]:    age  id     name  fingers languages
         0   47   1   dillon        9    python
         1   27   2     juan       10    python
```

Note that we have done an "inner join" here, which means we are only getting the intersection of the two tables. If we want the union, we can specify that we want an outer join:

```
In [34]: table.merge(other_table, on='name', how='outer')

Out[34]:    age  id     name  fingers languages
         0   47   1   dillon        9    python
         1   27   2     juan       10    python
         2   23   3   andrew      NaN       NaN
         3  NaN NaN      dav      NaN    python
```

Or maybe we want all of the data from `table`, but not `other_table`

```
In [35]: table.merge(other_table, on='name', how='left')

Out[35]:    age  id     name  fingers languages
         0   47   1   dillon        9    python
         1   27   2     juan       10    python
         2   23   3   andrew      NaN       NaN
```

2.3 Reshaping

To make analysis easier, you may have to reshape your data. It's easiest to deal with data when each table meets the follwing criteria:

1. Each row is exactly one observation
2. Each column is exactly one kind of data
3. The table expresses one and only one relationship between observations and variables

This kind of format is easy to work with, because:

1. It's easy to update when every piece of data exists in one and only one place
2. It's easy to subset conditionally across rows
3. It's easy to test across columns

To make this more concrete, let's take an example table.

name	city1	city2	population
dillon	williamsburg	berkeley	110
juan	berkeley	berkeley	110
dav	cambridge	berkeley	110

This table violates all three of the rules above. Specifically, it:

1. each row is about two observations
2. two columns are about the same kind of date (city), while another datatype (time) has been hidden in the column names
3. it expresses the relationship between people and where they live; and, cities and their population

In this particular example, our data is too wide. If we create that dataframe in pandas

```
In [36]: wide_table = pd.DataFrame({'name' : ['dillon', 'juan', 'dav'],
                                    'city1' : ['williamsburg', 'berkeley', 'cambridge'],
                                    'city2' : ['berkeley', 'berkeley', 'berkeley'],
                                    'population' : [110, 110, 110]
                                   })
         wide_table
```

```
Out[36]:         city1      city2    name  population
         0  williamsburg  berkeley  dillon         110
         1      berkeley  berkeley    juan         110
         2     cambridge  berkeley     dav         110
```

We can make this longer in pandas using the `melt` function

```
In [37]: long_table = pd.melt(wide_table, id_vars = ['name'])
         long_table
```

```
Out[37]:     name     variable         value
         0  dillon        city1  williamsburg
         1    juan        city1      berkeley
         2     dav        city1     cambridge
         3  dillon        city2      berkeley
         4    juan        city2      berkeley
         5     dav        city2      berkeley
         6  dillon   population           110
         7    juan   population           110
         8     dav   population           110
```

We can make the table wider using the pivot method

side note - this kind of inconsistency between `melt` and `pivot` is un-pythonic and should not be emulated

```
In [38]: long_table.pivot(columns='variable')
```

```
Out[38]:             name                              value
         variable   city1   city2 population        city1    city2 population
         0         dillon     NaN        NaN  williamsburg      NaN        NaN
         1           juan     NaN        NaN      berkeley      NaN        NaN
         2            dav     NaN        NaN     cambridge      NaN        NaN
         3            NaN  dillon        NaN           NaN  berkeley        NaN
         4            NaN    juan        NaN           NaN  berkeley        NaN
         5            NaN     dav        NaN           NaN  berkeley        NaN
         6            NaN     NaN     dillon           NaN      NaN        110
         7            NaN     NaN       juan           NaN      NaN        110
         8            NaN     NaN        dav           NaN      NaN        110
```

WHOA

One of the really cool things about pandas is that it allows you to have multiple indexes for rows and columns. Since pandas couldn't figure out what do with two kinds of value variables, it doubled up our column index. We can fix this by specifying that we only want the 'values' values

```
In [39]: long_table.pivot(index='name', columns='variable', values='value')

Out[39]: variable          city1      city2 population
         name
         dav            cambridge  berkeley        110
         dillon      williamsburg  berkeley        110
         juan            berkeley  berkeley        110
```

Challenge time! Switch computers *again* so that you are working on the first computer of the day, and have a look at challenge C. This will have you practice reading and merging tables. Again, when you are finished, check your work by running `py.test test_C` in a shell.

2.4 Descriptive statistics

Single descriptives have their own method calls in the `Series` class.

```
In [40]: table['fingers'].mean()

Out[40]: 9.5

In [41]: table['fingers'].std()

Out[41]: 0.70710678118654757

In [42]: table['fingers'].quantile(.25)

Out[42]: 9.25

In [43]: table['fingers'].kurtosis()

Out[43]: nan
```

You can call several of these at once with the `describe` method

```
In [44]: table.describe()

Out[44]:              age   id     fingers
         count   3.000000  3.0    2.000000
         mean   32.333333  2.0    9.500000
         std    12.858201  1.0    0.707107
         min    23.000000  1.0    9.000000
         25%    25.000000  1.5    9.250000
         50%    27.000000  2.0    9.500000
         75%    37.000000  2.5    9.750000
         max    47.000000  3.0   10.000000
```

2.5 Plotting

Sometimes, it can be easier to visualize results than to look at tables of numbers, both when you are sharing results and when you are exploring relationships in your data. To do this in Python, we'll be using `matplotlib`, the **Mat**lab **plot**ing **lib**rary.

```
In [45]: %matplotlib inline
         from matplotlib import pyplot as plt
```

To practice plotting, we'll pull in some real world data – the very same D-Lab feedback form data that you've been filling out after each day. (I told you we looked at it. I wasn't lying).

Most of the identifying information has been removed here, so what you'll see is better understood as feedback about the D-Lab in general, and not about a specific instructor or training. That said, some of the verbatims name individual instructors, but we ask that you not look into them individually.

```
In [46]: data = pd.read_csv('../data/03_feedback.csv')
         data.columns
```

```
Out[46]: Index(['timestamp', 'course_delivered', 'instructor_communicated', 'hear',
                'interest', 'department', 'verbs', 'useful', 'gender', 'ethnicity',
                'outside_barriers', 'inside_barriers', 'what_barriers', 'position'],
               dtype='object')
```

We have several kinds of data from each respondent, including their opinions about the world: > Do you experience barriers to access outside of the D-Lab (scale of 1-5)

Their experiences here: > Do you experience these barriers inside the D-Lab (scale of 1-5)

And some personal information, like their academic department and gender.

We can start looking at the distributions of some of these data with histograms. Plotting numerical data in `matplotlib` is fairly simple:

```
In [47]: plt.hist(data.outside_barriers, bins=5, range=[0,5])
         plt.show()
```
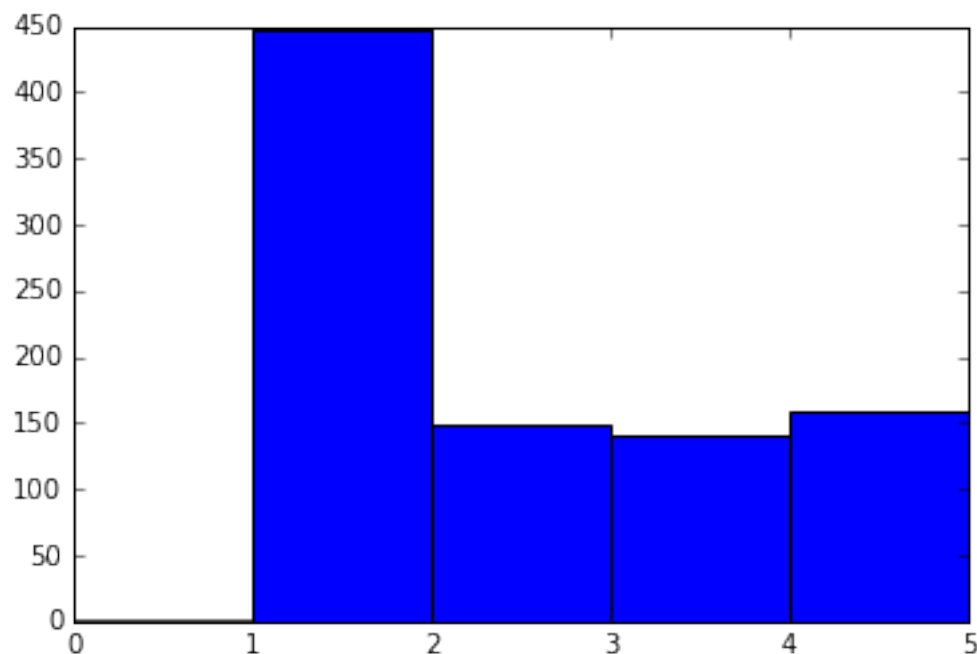

In `matplotlib`, there are no defaults around non-numeric data, including:

1. where to place things
2. how to name things

3. how to summarize things

so the first things we'll need to do are create iterables with x values, x labels, and y values

```
In [48]: index = range(len(data.gender.unique()))
         x_ticks = data.gender.unique()
         y_vals = [data[data.gender == value].outside_barriers.mean() for value in x_ticks]
         y_error = [data[data.gender == value].outside_barriers.std() for value in x_ticks]
         plt.bar(index, y_vals, width=0.8, color = 'r', yerr=y_error)
         plt.xticks([item + 0.4 for item in index], x_ticks, rotation=90)
         plt.show()
```
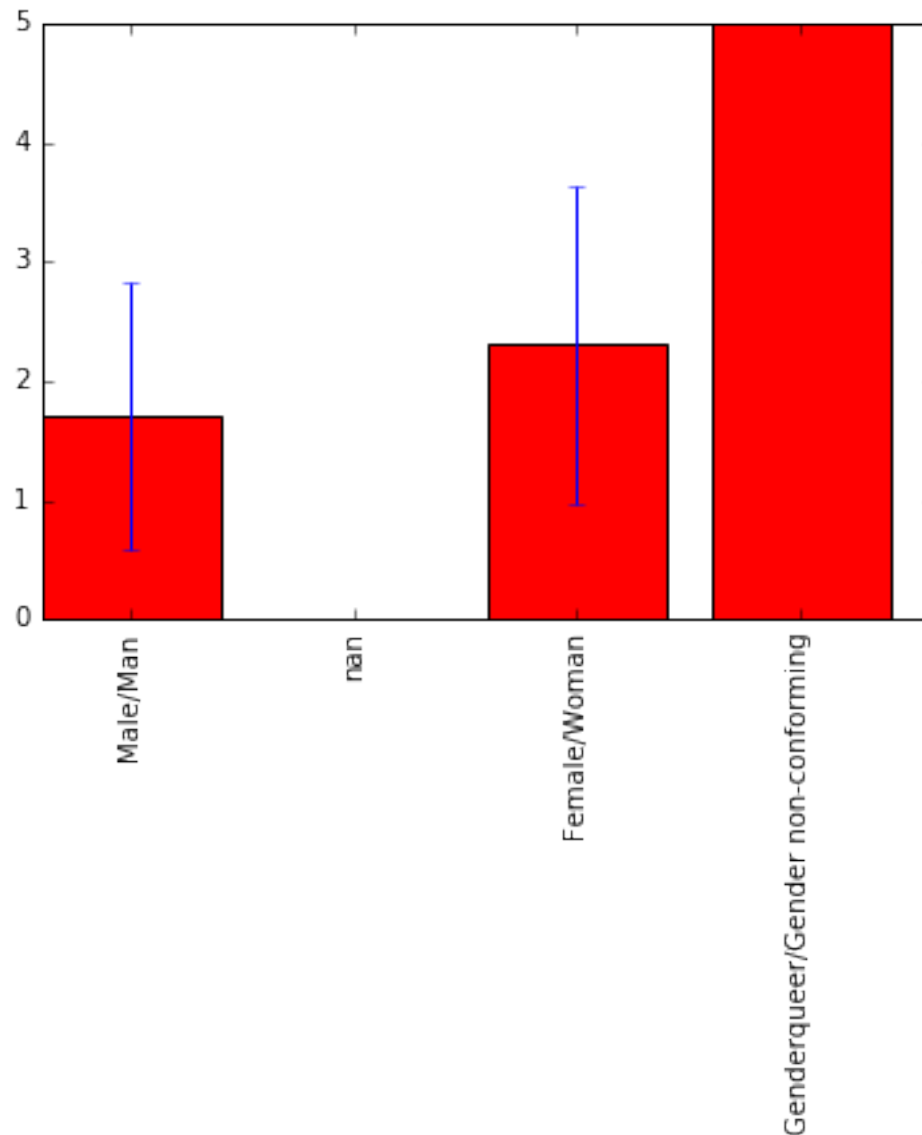
We can try plotting the values for inside barriers versus outside barriers, to see if it looks like there is any kind of relationship there, but we're going to run into a problem...

```
In [49]: plt.scatter(data.outside_barriers, data.inside_barriers)
         plt.show()
```
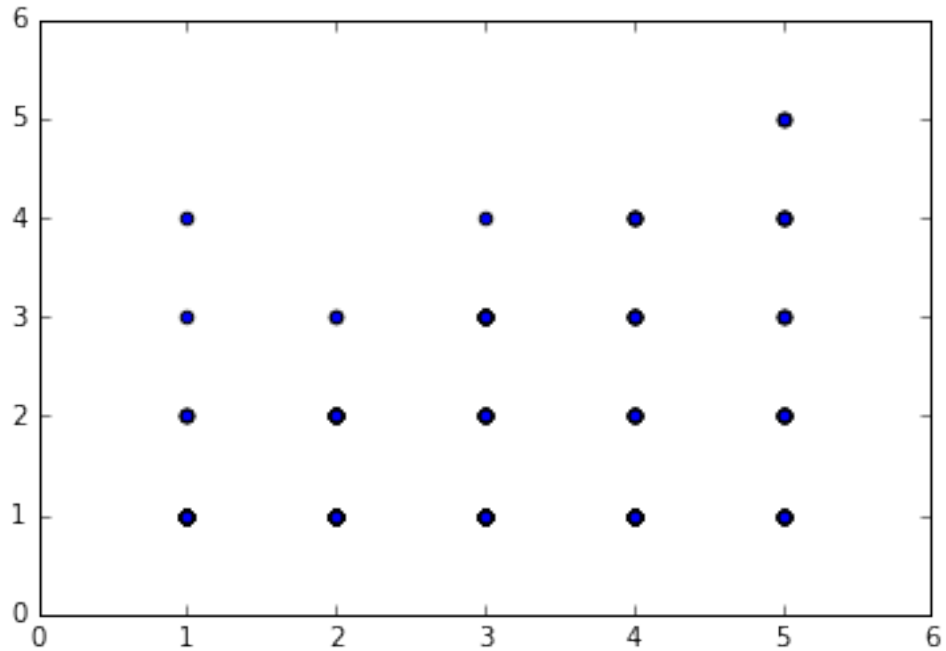
The problem that we have is that these are scales, with values like 1 and 2 - there are no floats like 1.189, so the dots *all overlap*.

We can fix this by adding a little bit of random noise to each value.

```
In [50]: from numpy.random import random
         x_vals = [value + random() for value in data.outside_barriers]
         y_vals = [value + random() for value in data.inside_barriers]
         plt.scatter(x_vals, y_vals)
         plt.show()
```
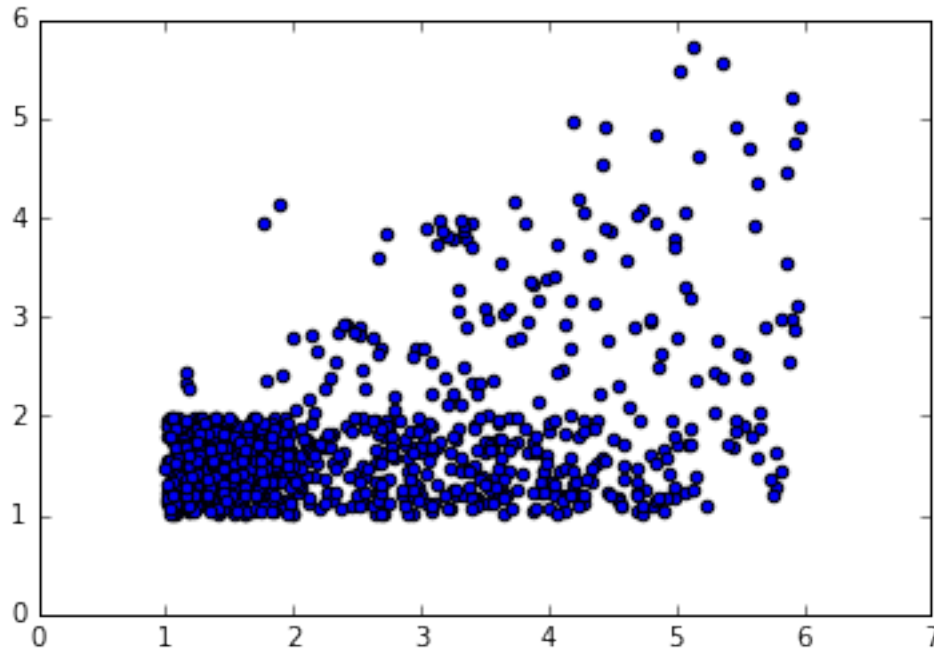
matplotlib is very customizable, but that means it also has a steep learning curve. If you'd like to know more, D-Lab does a whole workshop just on this one library.

2.6 Inferential statistics

pandas does not have statistical functions baked in, so we are going to call them from the `scipy.stats` library and the `statmodels` scikit.

We are also going to load in an actual dataset, as stats examples aren't very interesting with tiny bits of fake data.

```
In [51]: from scipy import stats
```

Using what you've learned so far about manipulating pandas objects, how would you find out the names of the variables in this dataset? Their datatypes? The distribution of their values?

2.6.1 Comparisons of group means

A common statistical procedure is to look for differences between groups of values. Typically, the values are grouped by a variable of interest, like sex or age. Here, we are going to compare the barriers of access to technology that people experience in the D-Lab compared to the world outside.

If you only have two groups in your sample, you can use a t-test:

```
In [52]: i = data['inside_barriers'].dropna()
         o = data['outside_barriers'].dropna()
         stats.ttest_ind(i, o)
```

```
Out[52]: Ttest_indResult(statistic=-16.595371177338013, pvalue=1.2776894527836828e-57)
```

Notice that here, we are passing in two whole columns, but we could also be subsetting by some other factor.

If you have more than two groups (or levels) that you would like to compare, you'll have to use something like an ANOVA:

```
In [53]: m = data[data.gender == "Male/Man"]['outside_barriers'].dropna()
         f = data[data.gender == "Female/Woman"]['outside_barriers'].dropna()
         q = data[data.gender == "Genderqueer/Gender non-conforming"]['outside_barriers'].dropna()
         stats.f_oneway(m, f, q)
```

```
Out[53]: F_onewayResult(statistic=24.849003218034316, pvalue=3.2400472234376748e-11)
```

Linear relationships Another common task is to establish if/how two variables are related across linear space. This could be something, for example, like relating shoe size to height. Here, we are going to ask whether barriers to access to technology inside and outside of the D-Lab are related.

One implementation of linear relationships is correlation testing:

```
In [54]: intermediate = data.dropna(subset=['inside_barriers', 'outside_barriers'])
         stats.pearsonr(intermediate['outside_barriers'], intermediate['inside_barriers'])
```

```
Out[54]: (0.46403957585300182, 2.075286653965493e-48)
```

At this point, we're going to pivot to using `statsmodels`

```
In [55]: import statsmodels.formula.api as smf
```

The formulas module in statsmodels lets us work with pandas dataframes, and linear model specifications that are similar to R and other variants of statistical software, e.g.:

```
outcome ~ var1 + var2
```

```
In [56]: model_1 = smf.ols("inside_barriers ~ outside_barriers", data=data).fit()
         model_1
```

```
Out[56]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x112b50eb8>
```

To get a summary of the test results, call the model's `summary` method

```
In [57]: model_1.summary()
```

```
Out[57]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                    OLS Regression Results
         ==============================================================================
         Dep. Variable:        inside_barriers   R-squared:                       0.215
         Model:                            OLS   Adj. R-squared:                  0.214
         Method:                 Least Squares   F-statistic:                     242.0
         Date:                Wed, 11 May 2016   Prob (F-statistic):           2.08e-48
         Time:                        14:51:38   Log-Likelihood:                 -807.74
         No. Observations:                 884   AIC:                             1619.
         Df Residuals:                     882   BIC:                             1629.
         Df Model:                           1
         Covariance Type:            nonrobust
         ==============================================================================
                             coef    std err          t      P>|t|      [95.0% Conf. Int.]
         ------------------------------------------------------------------------------
         Intercept         0.7529      0.038     19.599      0.000       0.678      0.828
         outside_barriers  0.2464      0.016     15.558      0.000       0.215      0.277
         ==============================================================================
         Omnibus:                      389.637   Durbin-Watson:                   1.865
         Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1871.839
         Skew:                           2.026   Prob(JB):                         0.00
```

```
        Kurtosis:                      8.865   Cond. No.                         5.17
        ====================================================================================
```

```
        Warnings:
        [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
        """
```

Since Python does not have private data or hidden attributes, you can pull out just about any intermediate information you want, including coefficients, residuals, and eigenvalues

Raymond Hettinger would say that Python is a "consenting adult language"

```
In [58]: model_1.params['outside_barriers']
```

```
Out[58]: 0.24638149915096616
```

statsmodels also exposes methods for validity checking your regressions, like looking for outliers by influence statistics

```
In [59]: model_1.get_influence().summary_frame()
```

```
Out[59]:      dfb_Intercept  dfb_outside_barriers      cooks_d     dffits  \
        0         -0.007772             0.000628  9.397542e-05 -0.013703
        1          0.000048            -0.000032  1.258195e-09  0.000050
        2          0.000048            -0.000032  1.258195e-09  0.000050
        3          0.002573            -0.020131  5.793345e-04 -0.034033
        4          0.000048            -0.000032  1.258195e-09  0.000050
        5          0.000048            -0.000032  1.258195e-09  0.000050
        6          0.000048            -0.000032  1.258195e-09  0.000050
        8          0.000048            -0.000032  1.258195e-09  0.000050
        9          0.000048            -0.000032  1.258195e-09  0.000050
        10         0.031181            -0.062463  2.801289e-03 -0.074872
        11         0.000048            -0.000032  1.258195e-09  0.000050
        12         0.000048            -0.000032  1.258195e-09  0.000050
        13         0.000048            -0.000032  1.258195e-09  0.000050
        14         0.000048            -0.000032  1.258195e-09  0.000050
        16         0.000048            -0.000032  1.258195e-09  0.000050
        18         0.002573            -0.020131  5.793345e-04 -0.034033
        20         0.002573            -0.020131  5.793345e-04 -0.034033
        21         0.000048            -0.000032  1.258195e-09  0.000050
        22         0.000048            -0.000032  1.258195e-09  0.000050
        23         0.031181            -0.062463  2.801289e-03 -0.074872
        24         0.002573            -0.020131  5.793345e-04 -0.034033
        26         0.000048            -0.000032  1.258195e-09  0.000050
        27         0.000048            -0.000032  1.258195e-09  0.000050
        28         0.023881            -0.001929  8.858263e-04  0.042104
        29         0.023881            -0.001929  8.858263e-04  0.042104
        30         0.031181            -0.062463  2.801289e-03 -0.074872
        32         0.031181            -0.062463  2.801289e-03 -0.074872
        33         0.000048            -0.000032  1.258195e-09  0.000050
        34        -0.007772             0.000628  9.397542e-05 -0.013703
        35         0.000048            -0.000032  1.258195e-09  0.000050
        ...             ...                  ...          ...       ...
        1032       0.000048            -0.000032  1.258195e-09  0.000050
        1033      -0.007772             0.000628  9.397542e-05 -0.013703
        1034       0.023881            -0.001929  8.858263e-04  0.042104
```

```
1035    0.000048              -0.000032  1.258195e-09  0.000050
1036    0.002573              -0.020131  5.793345e-04 -0.034033
1037    0.000048              -0.000032  1.258195e-09  0.000050
1038    0.000048              -0.000032  1.258195e-09  0.000050
1039   -0.007772               0.000628  9.397542e-05 -0.013703
1040    0.000048              -0.000032  1.258195e-09  0.000050
1041    0.000048              -0.000032  1.258195e-09  0.000050
1042    0.031181              -0.062463  2.801289e-03 -0.074872
1043    0.002573              -0.020131  5.793345e-04 -0.034033
1044   -0.002656               0.020780  6.172843e-04  0.035131
1045    0.000048              -0.000032  1.258195e-09  0.000050
1046   -0.007772               0.000628  9.397542e-05 -0.013703
1047    0.000048              -0.000032  1.258195e-09  0.000050
1048   -0.007772               0.000628  9.397542e-05 -0.013703
1049    0.000048              -0.000032  1.258195e-09  0.000050
1050    0.002573              -0.020131  5.793345e-04 -0.034033
1051    0.055758              -0.004503  4.791375e-03  0.098308
1052    0.000048              -0.000032  1.258195e-09  0.000050
1053    0.000048              -0.000032  1.258195e-09  0.000050
1054    0.000048              -0.000032  1.258195e-09  0.000050
1055    0.031181              -0.062463  2.801289e-03 -0.074872
1056    0.002573              -0.020131  5.793345e-04 -0.034033
1057    0.000048              -0.000032  1.258195e-09  0.000050
1058    0.031181              -0.062463  2.801289e-03 -0.074872
1059   -0.007772               0.000628  9.397542e-05 -0.013703
1060    0.000048              -0.000032  1.258195e-09  0.000050
1061    0.000048              -0.000032  1.258195e-09  0.000050

      dffits_internal  hat_diag  standard_resid  student_resid
0          -0.013710  0.001134       -0.406955      -0.406762
1           0.000050  0.001902        0.001149       0.001149
2           0.000050  0.001902        0.001149       0.001149
3          -0.034039  0.001740       -0.815305      -0.815150
4           0.000050  0.001902        0.001149       0.001149
5           0.000050  0.001902        0.001149       0.001149
6           0.000050  0.001902        0.001149       0.001149
8           0.000050  0.001902        0.001149       0.001149
9           0.000050  0.001902        0.001149       0.001149
10         -0.074850  0.003721       -1.224749      -1.225096
11          0.000050  0.001902        0.001149       0.001149
12          0.000050  0.001902        0.001149       0.001149
13          0.000050  0.001902        0.001149       0.001149
14          0.000050  0.001902        0.001149       0.001149
16          0.000050  0.001902        0.001149       0.001149
18         -0.034039  0.001740       -0.815305      -0.815150
20         -0.034039  0.001740       -0.815305      -0.815150
21          0.000050  0.001902        0.001149       0.001149
22          0.000050  0.001902        0.001149       0.001149
23         -0.074850  0.003721       -1.224749      -1.225096
24         -0.034039  0.001740       -0.815305      -0.815150
26          0.000050  0.001902        0.001149       0.001149
27          0.000050  0.001902        0.001149       0.001149
28          0.042091  0.001134        1.249433       1.249831
29          0.042091  0.001134        1.249433       1.249831
```

```
30         -0.074850   0.003721      -1.224749     -1.225096
32         -0.074850   0.003721      -1.224749     -1.225096
33          0.000050   0.001902       0.001149      0.001149
34         -0.013710   0.001134      -0.406955     -0.406762
35          0.000050   0.001902       0.001149      0.001149
...              ...        ...            ...           ...
1032        0.000050   0.001902       0.001149      0.001149
1033       -0.013710   0.001134      -0.406955     -0.406762
1034        0.042091   0.001134       1.249433      1.249831
1035        0.000050   0.001902       0.001149      0.001149
1036       -0.034039   0.001740      -0.815305     -0.815150
1037        0.000050   0.001902       0.001149      0.001149
1038        0.000050   0.001902       0.001149      0.001149
1039       -0.013710   0.001134      -0.406955     -0.406762
1040        0.000050   0.001902       0.001149      0.001149
1041        0.000050   0.001902       0.001149      0.001149
1042       -0.074850   0.003721      -1.224749     -1.225096
1043       -0.034039   0.001740      -0.815305     -0.815150
1044        0.035136   0.001740       0.841585      0.841446
1045        0.000050   0.001902       0.001149      0.001149
1046       -0.013710   0.001134      -0.406955     -0.406762
1047        0.000050   0.001902       0.001149      0.001149
1048       -0.013710   0.001134      -0.406955     -0.406762
1049        0.000050   0.001902       0.001149      0.001149
1050       -0.034039   0.001740      -0.815305     -0.815150
1051        0.097892   0.001134       2.905821      2.918175
1052        0.000050   0.001902       0.001149      0.001149
1053        0.000050   0.001902       0.001149      0.001149
1054        0.000050   0.001902       0.001149      0.001149
1055       -0.074850   0.003721      -1.224749     -1.225096
1056       -0.034039   0.001740      -0.815305     -0.815150
1057        0.000050   0.001902       0.001149      0.001149
1058       -0.074850   0.003721      -1.224749     -1.225096
1059       -0.013710   0.001134      -0.406955     -0.406762
1060        0.000050   0.001902       0.001149      0.001149
1061        0.000050   0.001902       0.001149      0.001149

[884 rows x 8 columns]
```

If, at this stage, you suspect that one or more outliers is unduly influencing your model fit, you can transform your results into robust OLS with a method call:

In [60]: model_1.get_robustcov_results().summary()

Out[60]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                 OLS Regression Results
         ==
         Dep. Variable: inside_barriers   R-squared: 0.215
         Model: OLS   Adj. R-squared: 0.214
         Method: Least Squares   F-statistic: 101.0
         Date: Wed, 11 May 2016   Prob (F-statistic):   1.40e-22
         Time: 14:51:39   Log-Likelihood: -807.74
         No. Observations: 884   AIC: 1619.
         Df Residuals: 882   BIC: 1629.

```
Df Model:                              1
Covariance Type:                     HC1
==============================================================================
                    coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept         0.7529      0.036     21.041      0.000       0.683      0.823
outside_barriers  0.2464      0.025     10.052      0.000       0.198      0.294
==============================================================================
Omnibus:                      389.637   Durbin-Watson:                   1.865
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1871.839
Skew:                           2.026   Prob(JB):                         0.00
Kurtosis:                       8.865   Cond. No.                         5.17
==============================================================================

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)
"""
```

This isn't very different, so we're probably okay.

If you want to add more predictors to your model, you can do so inside the function string:

In [61]: `smf.ols("inside_barriers ~ outside_barriers + gender", data=data).fit().summary()`

Out[61]: `<class 'statsmodels.iolib.summary.Summary'>`
```
"""
                          OLS Regression Results
==============================================================================
Dep. Variable:         inside_barriers   R-squared:                       0.235
Model:                             OLS   Adj. R-squared:                  0.233
Method:                  Least Squares   F-statistic:                     86.65
Date:                 Wed, 11 May 2016   Prob (F-statistic):           7.02e-49
Time:                         14:51:39   Log-Likelihood:                 -748.48
No. Observations:                  848   AIC:                             1505.
Df Residuals:                      844   BIC:                             1524.
Df Model:                            3
Covariance Type:             nonrobust
=================================================================================
                                            coef    std err          t      P>|t|      [
---------------------------------------------------------------------------------
Intercept                                 0.6550      0.045     14.680      0.000
gender[T.Genderqueer/Gender non-conforming] -0.9480    0.588     -1.611      0.108
gender[T.Male/Man]                        0.1808      0.043      4.209      0.000
outside_barriers                          0.2586      0.016     16.106      0.000
=================================================================================
Omnibus:                      358.447   Durbin-Watson:                   1.947
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1650.675
Skew:                           1.940   Prob(JB):                         0.00
Kurtosis:                       8.626   Cond. No.                         76.2
=================================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
"""
```

Note that our categorical/factor variable has been automatically one-hot encoded as treatment conditions.

There's not way to change this within `statsmodels`, but you can specify your contrasts indirectly using a library called (`Patsy`)[http://statsmodels.sourceforge.net/stable/contrasts.html].

To add interactions to your model, you can use :, or * [for full factorial]

```
In [62]: smf.ols("inside_barriers ~ outside_barriers * gender", data=data).fit().summary()

Out[62]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                    OLS Regression Results
         ==============================================================================
         Dep. Variable:          inside_barriers   R-squared:                       0.267
         Model:                              OLS   Adj. R-squared:                  0.264
         Method:                   Least Squares   F-statistic:                     76.92
         Date:                  Wed, 11 May 2016   Prob (F-statistic):           1.26e-55
         Time:                          14:51:39   Log-Likelihood:                 -730.40
         No. Observations:                   848   AIC:                             1471.
         Df Residuals:                       843   BIC:                             1495.
         Df Model:                             4
         Covariance Type:               nonrobust
         ==============================================================================
                                                                   coef    std err          t
         ------------------------------------------------------------------------------------
         Intercept                                               0.7909      0.049     16.101
         gender[T.Genderqueer/Gender non-conforming]            -0.0303      0.022     -1.364
         gender[T.Male/Man]                                     -0.2132      0.077     -2.753
         outside_barriers                                        0.1993      0.019     10.755
         outside_barriers:gender[T.Genderqueer/Gender non-conforming]  -0.1514      0.111     -1.364
         outside_barriers:gender[T.Male/Man]                     0.2124      0.035      6.061
         ==============================================================================
         Omnibus:                        351.976   Durbin-Watson:                   1.962
         Prob(Omnibus):                    0.000   Jarque-Bera (JB):             1805.642
         Skew:                             1.851   Prob(JB):                         0.00
         Kurtosis:                         9.115   Cond. No.                     2.17e+16
         ==============================================================================

         Warnings:
         [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
         [2] The smallest eigenvalue is 1.31e-29. This might indicate that there are
         strong multicollinearity problems or that the design matrix is singular.
         """
```

3 Practice

In the time remaining, pull up a dataset that you have, and that you'd like to work with in Python. The instructors will be around to help you apply what you've learned today to problems in your data that you are dealing with.

If you don't have data of your own, you should practice with the test data we've given you here. For example, you could try to figure out:

1. Do different departments have the same gender ratios?
2. What variable in this dataset is the best predictor for how useful people find our workshops to be?

```
In [63]:
```