Preliminary Analysis

(1) Read the JSON file to Start Preliminary Analysis

```
In [1]:    import json
           from collections import Counter
           from pprint import pprint
```

```
In [2]:    infile = open('tweet_stream_COVID_200.json')
           data = json.load(infile)
           infile.close()
```

(2) Let's Investigate the Data Structure

```
In [3]:    type(data)
```

Out[3]:   list

```
In [4]:    len(data) # Each tweet is represented by an element in a list
```

Out[4]:   200

```
In [5]:    type(data[0]) # Data type of each tweet is a dictionary
```

Out[5]:   dict

```
In [6]:    data[0].keys()
```

Out[6]:   dict_keys(['created_at', 'id', 'id_str', 'text', 'source', 'truncated', 'in_reply_to_sta
          tus_id', 'in_reply_to_status_id_str', 'in_reply_to_user_id', 'in_reply_to_user_id_str',
          'in_reply_to_screen_name', 'user', 'geo', 'coordinates', 'place', 'contributors', 'retwe
          eted_status', 'quoted_status_id', 'quoted_status_id_str', 'quoted_status', 'quoted_statu
          s_permalink', 'is_quote_status', 'quote_count', 'reply_count', 'retweet_count', 'favorit
          e_count', 'entities', 'favorited', 'retweeted', 'filter_level', 'lang', 'timestamp_ms'])

```
In [7]:    pprint(data[0])
```

```
{'contributors': None,
 'coordinates': None,
 'created_at': 'Sat Oct 10 23:31:53 +0000 2020',
 'entities': {'hashtags': [],
              'symbols': [],
              'urls': [],
              'user_mentions': [{'id': 2949261587,
                                  'id_str': '2949261587',
                                  'indices': [3, 18],
                                  'name': 'Benjamin Dixon',
                                  'screen_name': 'BenjaminPDixon'}]},
```

 'favorite_count': 0,
 'favorited': False,
 'filter_level': 'low',
 'geo': None,
 'id': 13150725592051462145,
 'id_str': '13150725592051462145',
 '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': True,
 'lang': 'en',
 'place': None,
 'quote_count': 0,
 'quoted_status': {'contributors': None,
                   'coordinates': None,
                   'created_at': 'Sat Oct 10 15:28:05 +0000 2020',
                   'display_text_range': [0, 86],
                   'entities': {'hashtags': [],
                                'media': [{'display_url': 'pic.twitter.com/LBAdxCLW3c',
                                           'expanded_url': 'https://twitter.com/CNNPolit
ics/status/1314950838897569794/photo/1',
                                           'id': 1314950837073072128,
                                           'id_str': '1314950837073072128',
                                           'indices': [87, 110],
                                           'media_url': 'http://pbs.twimg.com/media/Ej-k
0cTX0AAKRW-.jpg',
                                           'media_url_https': 'https://pbs.twimg.com/med
ia/Ej-k0cTX0AAKRW-.jpg',
                                           'sizes': {'large': {'h': 619,
                                                               'resize': 'fit',
                                                               'w': 1100},
                                                     'medium': {'h': 619,
                                                                'resize': 'fit',
                                                                'w': 1100},
                                                     'small': {'h': 383,
                                                               'resize': 'fit',
                                                               'w': 680},
                                                     'thumb': {'h': 150,
                                                               'resize': 'crop',
                                                               'w': 150}},
                                           'type': 'photo',
                                           'url': 'https://t.co/LBAdxCLW3c'}],
                                'symbols': [],
                                'urls': [{'display_url': 'cnn.it/36RbcJp',
                                          'expanded_url': 'https://cnn.it/36RbcJp',
                                          'indices': [63, 86],
                                          'url': 'https://t.co/paIuNK7Sgp'}],
                                'user_mentions': []},
                   'extended_entities': {'media': [{'display_url': 'pic.twitter.com/LBAd
xCLW3c',
                                                    'expanded_url': 'https://twitter.co
m/CNNPolitics/status/1314950838897569794/photo/1',
                                                    'id': 1314950837073072128,
                                                    'id_str': '1314950837073072128',
                                                    'indices': [87, 110],
                                                    'media_url': 'http://pbs.twimg.com/m
edia/Ej-k0cTX0AAKRW-.jpg',
                                                    'media_url_https': 'https://pbs.twim
g.com/media/Ej-k0cTX0AAKRW-.jpg',
                                                    'sizes': {'large': {'h': 619,
                                                                        'resize': 'fit',
                                                                        'w': 1100},
                                                              'medium': {'h': 619,

                                                                        'resize': 'fi
t',
                                                                        'w': 1100},
                                                    'small': {'h': 383,
                                                              'resize': 'fit',
                                                              'w': 680},
                                                    'thumb': {'h': 150,
                                                              'resize': 'cro
p',
                                                              'w': 150}},
                                        'type': 'photo',
                                        'url': 'https://t.co/LBAdxCLW3c'}]},
              'favorite_count': 516,
              'favorited': False,
              'filter_level': 'low',
              'geo': None,
              'id': 1314950838897569794,
              'id_str': '1314950838897569794',
              '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',
              'place': None,
              'possibly_sensitive': False,
              'quote_count': 183,
              'reply_count': 158,
              'retweet_count': 107,
              'retweeted': False,
              'source': '<a href="http://www.socialflow.com" '
                        'rel="nofollow">SocialFlow',
              'text': 'Chris Christie released from hospital after '
                      'Covid-19 treatment https://t.co/paIuNK7Sgp '
                      'https://t.co/LBAdxCLW3c',
              'truncated': False,
              'user': {'contributors_enabled': False,
                       'created_at': 'Sat Feb 23 03:12:49 +0000 2008',
                       'default_profile': False,
                       'default_profile_image': False,
                       'description': 'Political news, campaign stories '
                                      'and Washington coverage from CNN '
                                      'Politics.',
                       'favourites_count': 4,
                       'follow_request_sent': None,
                       'followers_count': 3726464,
                       'following': None,
                       'friends_count': 344,
                       'geo_enabled': False,
                       'id': 13850422,
                       'id_str': '13850422',
                       'is_translator': False,
                       'lang': None,
                       'listed_count': 18153,
                       'location': 'Washington, DC',
                       'name': 'CNN Politics',
                       'notifications': None,
                       'profile_background_color': '100F14',
                       'profile_background_image_url': 'http://abs.twimg.com/image
s/themes/theme1/bg.png',
                       'profile_background_image_url_https': 'https://abs.twimg.co
m/images/themes/theme1/bg.png',
                       'profile_background_tile': True,
                       'profile_banner_url': 'https://pbs.twimg.com/profile_banner

```
                    s/13850422/1503504077',
                                'profile_image_url': 'http://pbs.twimg.com/profile_images/91
                    88899077168934912/NrRRE0_b_normal.jpg',
                                'profile_image_url_https': 'https://pbs.twimg.com/profile_im
                    ages/918899077168934912/NrRRE0_b_normal.jpg',
                                'profile_link_color': '0000FF',
                                'profile_sidebar_border_color': 'FFFFFF',
                                'profile_sidebar_fill_color': 'F1F5E7',
                                'profile_text_color': '000000',
                                'profile_use_background_image': True,
                                'protected': False,
                                'screen_name': 'CNNPolitics',
                                'statuses_count': 188997,
                                'time_zone': None,
                                'translator_type': 'none',
                                'url': 'http://cnn.com/politics',
                                'utc_offset': None,
                                'verified': True}},
 'quoted_status_id': 1314950838897569794,
 'quoted_status_id_str': '1314950838897569794',
 'quoted_status_permalink': {'display': 'twitter.com/cnnpolitics/st…',
                             'expanded': 'https://twitter.com/cnnpolitics/status/1314950
                    838897569794',
                             'url': 'https://t.co/kkYrdTk1Cv'},
 'reply_count': 0,
 'retweet_count': 0,
 'retweeted': False,
 'retweeted_status': {'contributors': None,
                      'coordinates': None,
                      'created_at': 'Sat Oct 10 15:30:55 +0000 2020',
                      'entities': {'hashtags': [],
                                   'symbols': [],
                                   'urls': [{'display_url': 'twitter.com/i/web/status/
                    1…',
                                             'expanded_url': 'https://twitter.com/i/web/
                    status/1314951552499683329',
                                             'indices': [117, 140],
                                             'url': 'https://t.co/MKXVYHrir7'}],
                                   'user_mentions': []},
                      'extended_tweet': {'display_text_range': [0, 185],
                                         'entities': {'hashtags': [],
                                                      'symbols': [],
                                                      'urls': [],
                                                      'user_mentions': []},
                                         'full_text': 'Just saying... if '
                                                      'Christie and Trump '
                                                      'survive while '
                                                      'hundreds-of-thousands '
                                                      'are dying, we're gonna '
                                                      'have to have an '
                                                      'in-person discussion '
                                                      'about what meds they '
                                                      'have that millions of '
                                                      'us don't.'},
                      'favorite_count': 39659,
                      'favorited': False,
                      'filter_level': 'low',
                      'geo': None,
                      'id': 1314951552499683329,
                      'id_str': '1314951552499683329',
                      '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': True,
                              'lang': 'en',
                              'place': None,
                              'quote_count': 151,
                              'quoted_status': {'contributors': None,
                                                'coordinates': None,
                                                'created_at': 'Sat Oct 10 15:28:05 '
                                                              '+0000 2020',
                                                'display_text_range': [0, 86],
                                                'entities': {'hashtags': [],
                                                             'media': [{'display_url': 'pic.twit
ter.com/LBAdxCLW3c',
                                                                        'expanded_url': 'http
s://twitter.com/CNNPolitics/status/1314950838897569794/photo/1',
                                                                        'id': 131495083707307212
8,
                                                                        'id_str': '1314950837073
072128',
                                                                        'indices': [87,
                                                                                    110],
                                                                        'media_url': 'http://pb
s.twimg.com/media/Ej-k0cTX0AAKRW-.jpg',
                                                                        'media_url_https': 'http
s://pbs.twimg.com/media/Ej-k0cTX0AAKRW-.jpg',
                                                                        'sizes': {'large': {'h':
619,
                                                                                            'res
ize': 'fit',
                                                                                            'w':
1100},
                                                                                  'medium':
{'h': 619,
                                                                                             're
size': 'fit',
                                                                                             'w': 1100},
                                                                                  'small': {'h':
383,
                                                                                            'res
ize': 'fit',
                                                                                            'w':
680},
                                                                                  'thumb': {'h':
150,
                                                                                            'res
ize': 'crop',
                                                                                            'w':
150}},
                                                                        'type': 'photo',
                                                                        'url': 'https://t.co/LBA
dxCLW3c'}],
                                                             'symbols': [],
                                                             'urls': [{'display_url': 'cnn.it/36
RbcJp',
                                                                       'expanded_url': 'https://
cnn.it/36RbcJp',
                                                                       'indices': [63,
                                                                                   86],
                                                                       'url': 'https://t.co/paIu
NK7Sgp'}],
                                                             'user_mentions': []},
                                                'extended_entities': {'media': [{'display_url':
'pic.twitter.com/LBAdxCLW3c',
                                                                                 'expanded_url':
'https://twitter.com/CNNPolitics/status/1314950838897569794/photo/1',

7073072128,

950837073072128',

0],

ttp://pbs.twimg.com/media/Ej-k0cTX0AAKRW-.jpg',

s': 'https://pbs.twimg.com/media/Ej-k0cTX0AAKRW-.jpg',

e': {'h': 619,

'resize': 'fit',

'w': 1100},

um': {'h': 619,

'resize': 'fit',

'w': 1100},

l': {'h': 383,

'resize': 'fit',

'w': 680},

b': {'h': 150,

'resize': 'crop',

'w': 150}},

o',

s://t.co/LBAdxCLW3c'}]},

'id': 131495083

'id_str': '1314

'indices': [87,
                11

'media_url': 'h

'media_url_http

'sizes': {'larg

                        'medi

                        'smal

                        'thum

'type': 'phot

'url': 'http

'favorite_count': 516,
'favorited': False,
'filter_level': 'low',
'geo': None,
'id': 1314950838897569794,
'id_str': '1314950838897569794',
'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',
'place': None,
'possibly_sensitive': False,
'quote_count': 183,
'reply_count': 158,
'retweet_count': 107,
'retweeted': False,
'source': '<a '
          'href="http://www.socialflow.com" '
          'rel="nofollow">SocialFlow',
'text': 'Chris Christie released from '
        'hospital after Covid-19 '
        'treatment '
        'https://t.co/paIuNK7Sgp '

                                                      'https://t.co/LBAdxCLW3c',
                                        'truncated': False,
                                        'user': {'contributors_enabled': False,
                                                 'created_at': 'Sat Feb 23 '
                                                               '03:12:49 +0000 '
                                                               '2008',
                                                 'default_profile': False,
                                                 'default_profile_image': False,
                                                 'description': 'Political '
                                                                'news, '
                                                                'campaign '
                                                                'stories and '
                                                                'Washington '
                                                                'coverage from '
                                                                'CNN Politics.',
                                                 'favourites_count': 4,
                                                 'follow_request_sent': None,
                                                 'followers_count': 3726464,
                                                 'following': None,
                                                 'friends_count': 344,
                                                 'geo_enabled': False,
                                                 'id': 13850422,
                                                 'id_str': '13850422',
                                                 'is_translator': False,
                                                 'lang': None,
                                                 'listed_count': 18153,
                                                 'location': 'Washington, DC',
                                                 'name': 'CNN Politics',
                                                 'notifications': None,
                                                 'profile_background_color': '100F14',
                                                 'profile_background_image_url': 'htt
p://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': True,
                                                 'profile_banner_url': 'https://pbs.twim
g.com/profile_banners/13850422/1503504077',
                                                 'profile_image_url': 'http://pbs.twimg.
com/profile_images/918899077168934912/NrRRE0_b_normal.jpg',
                                                 'profile_image_url_https': 'https://pb
s.twimg.com/profile_images/918899077168934912/NrRRE0_b_normal.jpg',
                                                 'profile_link_color': '0000FF',
                                                 'profile_sidebar_border_color': 'FFFFF
F',
                                                 'profile_sidebar_fill_color': 'F1F5E7',
                                                 'profile_text_color': '000000',
                                                 'profile_use_background_image': True,
                                                 'protected': False,
                                                 'screen_name': 'CNNPolitics',
                                                 'statuses_count': 188997,
                                                 'time_zone': None,
                                                 'translator_type': 'none',
                                                 'url': 'http://cnn.com/politics',
                                                 'utc_offset': None,
                                                 'verified': True}},
                              'quoted_status_id': 1314950838897569794,
                              'quoted_status_id_str': '1314950838897569794',
                              'quoted_status_permalink': {'display': 'twitter.com/cnnpolitics/s
t…',
                                                          'expanded': 'https://twitter.com/cnnpo
litics/status/1314950838897569794',
                                                          'url': 'https://t.co/kkYrdTk1Cv'},
                              'reply_count': 298,
                              'retweet_count': 8530,
                              'retweeted': False,

                              'source': '<a href="http://twitter.com/download/iphone" '
                                        'rel="nofollow">Twitter for iPhone',
                              'text': 'Just saying... if Christie and Trump survive '
                                      'while hundreds-of-thousands are dying, we're '
                                      'gonna have to have an in-… '
                                      'https://t.co/MKXVYHrir7',
                              'truncated': True,
                              'user': {'contributors_enabled': False,
                                       'created_at': 'Mon Dec 29 13:58:38 +0000 2014',
                                       'default_profile': False,
                                       'default_profile_image': False,
                                       'description': '"when I opened my eyes and saw '
                                                      'the real world, I began to '
                                                      'laugh and I haven't stopped '
                                                      'since."',
                                       'favourites_count': 182600,
                                       'follow_request_sent': None,
                                       'followers_count': 183826,
                                       'following': None,
                                       'friends_count': 4164,
                                       'geo_enabled': True,
                                       'id': 2949261587,
                                       'id_str': '2949261587',
                                       'is_translator': False,
                                       'lang': None,
                                       'listed_count': 1263,
                                       'location': 'Boston',
                                       'name': 'Benjamin Dixon',
                                       'notifications': None,
                                       'profile_background_color': '000000',
                                       'profile_background_image_url': 'http://abs.twimg.com/ima
ges/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_bann
ers/2949261587/1593480554',
                                       'profile_image_url': 'http://pbs.twimg.com/profile_image
s/1151836172148203520/xSn697lX_normal.jpg',
                                       'profile_image_url_https': 'https://pbs.twimg.com/profile
_images/1151836172148203520/xSn697lX_normal.jpg',
                                       'profile_link_color': '460067',
                                       'profile_sidebar_border_color': '000000',
                                       'profile_sidebar_fill_color': '000000',
                                       'profile_text_color': '000000',
                                       'profile_use_background_image': False,
                                       'protected': False,
                                       'screen_name': 'BenjaminPDixon',
                                       'statuses_count': 88086,
                                       'time_zone': None,
                                       'translator_type': 'none',
                                       'url': 'http://youtube.com/thebenjamindixonshow',
                                       'utc_offset': None,
                                       'verified': False}},
  'source': '<a href="http://twitter.com/download/iphone" '
            'rel="nofollow">Twitter for iPhone',
  'text': 'RT @BenjaminPDixon: Just saying... if Christie and Trump survive '
          'while hundreds-of-thousands are dying, we're gonna have to have an '
          'in-pers…',
  'timestamp_ms': '1602372713469',
  'truncated': False,
  'user': {'contributors_enabled': False,
           'created_at': 'Sun Aug 12 16:42:20 +0000 2012',
           'default_profile': False,
           'default_profile_image': False,

```
          'description': None,
          'favourites_count': 20035,
          'follow_request_sent': None,
          'followers_count': 105,
          'following': None,
          'friends_count': 499,
          'geo_enabled': False,
          'id': 753407732,
          'id_str': '753407732',
          'is_translator': False,
          'lang': None,
          'listed_count': 2,
          'location': None,
          'name': 'J Donovan',
          'notifications': None,
          '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/the
me1/bg.png',
          'profile_background_tile': True,
          'profile_banner_url': 'https://pbs.twimg.com/profile_banners/753407732/1355965
550',
          'profile_image_url': 'http://pbs.twimg.com/profile_images/931972757885149184/n
bDfZg-N_normal.jpg',
          'profile_image_url_https': 'https://pbs.twimg.com/profile_images/9319727578851
49184/nbDfZg-N_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': 'JadenSai',
          'statuses_count': 18050,
          'time_zone': None,
          'translator_type': 'none',
          'url': None,
          'utc_offset': None,
          'verified': False}}
```

In [8]:
```python
data[0]['entities']
```

Out[8]:
```
{'hashtags': [],
 'urls': [],
 'user_mentions': [{'screen_name': 'BenjaminPDixon',
   'name': 'Benjamin Dixon',
   'id': 2949261587,
   'id_str': '2949261587',
   'indices': [3, 18]}],
 'symbols': []}
```

In [9]:
```python
for i in data:
    if i['entities']['hashtags'] != []:
        print(data.index(i))
```

```
18
26
45
47
48
56
80
```

```
112
152
161
184
```

In [10]:
```python
data[18]['entities']['hashtags']
```

Out[10]:
```
[{'text': 'Canada', 'indices': [20, 27]},
 {'text': 'COVID19', 'indices': [55, 63]},
 {'text': 'Alberta', 'indices': [116, 124]},
 {'text': 'Quebec', 'indices': [126, 133]}]
```

In [11]:
```python
data[18]['entities']['hashtags'][0]['text']
```

Out[11]:
```
'Canada'
```

In [12]:
```python
for t in data:
    for h in t['entities']['hashtags']:
        print(h['text'])
```

```
Canada
COVID19
Alberta
Quebec
covid19
YouShouldBeHere
Cuba
Miami
USA
DemVoice1
ElCovidMata
LGBTQ
Ahora
coronavirus
BorisJohnson
MattHancock
Cineworld
TESTIMONIO
```

In [13]:
```python
data[0]['user']
```

Out[13]:
```
{'id': 753407732,
 'id_str': '753407732',
 'name': 'J Donovan',
 'screen_name': 'JadenSai',
 'location': None,
 'url': None,
 'description': None,
 'translator_type': 'none',
 'protected': False,
 'verified': False,
 'followers_count': 105,
 'friends_count': 499,
 'listed_count': 2,
 'favourites_count': 20035,
 'statuses_count': 18050,
 'created_at': 'Sun Aug 12 16:42:20 +0000 2012',
 'utc_offset': None,
 'time_zone': None,
```

```
 'geo_enabled': False,
 'lang': None,
 'contributors_enabled': False,
 'is_translator': 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': True,
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'C0DEED',
 'profile_sidebar_fill_color': 'DDEEF6',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'profile_image_url': 'http://pbs.twimg.com/profile_images/931972757885149184/nbDfZg-N_normal.jpg',
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/931972757885149184/nbDfZg-N_normal.jpg',
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/753407732/1355965550',
 'default_profile': False,
 'default_profile_image': False,
 'following': None,
 'follow_request_sent': None,
 'notifications': None}
```

In [14]:
```python
data[0]['user']['name']
```

Out[14]: 'J Donovan'

In [15]:
```python
for t in data:
    print(t['user']['name'])
```

```
J Donovan
jb
Canadian patriot
BarksBeachesBooks_Kate
D
lastfighter
Shadow
Fiti
Mateo
DiMusa
Sharon Schatz
quack
KS
LALaN🥇🥇健康でいてください⛏🍄⛏🍄
aloha_kids2
Ship_Love
smilealice12
Maxi Buteler
A Fair Canada
Jim E-H
Trace
jamaica
Chy
el unico bolivar
Char74Trump
Dylan-149
Morgan Fitzpatrick, MPH, RN, NP
weary_ant
Kari Blizzard
Open Eugene
```

UnsilentDeaf 🇺🇸
Heerak Christian Kim for US Congress (Virginia-8)
Renita
Zara 🐾
Mary-Ann Noyes
Windycity
Benjamin Breece
Donna 🌍👩🦊🇺🇸
🇺🇸 Average Deplorable 🇺🇸
🆘Biden / Harris 2020🆘
eleanor, 🌷
jade nacht
Kristin Maier
S/o Mlk
Diogenes of 白左疆
Murphy Payeur
Sunday World
Capitan Manuel
Candice Phelps
The Reporters
7 estrellas
cosima van elgin
Renton
Dan Johno
JJ Gills
Roci🌻
Marilyn
Dave Mattingly
Tim
Ms. Banks ❤️
LLB
Carlos Toraño
BabyBoomer1
Matt Aronson
Oh for crying out loud! (I am totally a bot)
Jobsen20
Robots in Kashmir
Stefania8
messi
Tina Wilson
Emma - AHA+A+L
meɪr☆약속¹³♥
Feisty Female
Donna Barbieri
Jamie Kennedy 🏴🏳️UMUM
FREEPREETY
DeeDeeTee
Andreina
Amriana
SmilingAssassin27
NETO OLIVEIRA 🌻
Richard M Garner
Swapnil
Beef Queef in YoSelf
James
Jolly BLM Rogers
KGarrison
Russell 💣
KPM
Dr. Allee
Infectious Diseases
Christian Rosu
🪁Readers for Pete👟

Aii Ann Kei #RakyatJagaRakyat
Jamie Mayerfeld
Denia
Sadmasist 💀💀
TiaLolo
Víctor Manuel
Shari Sanders
Ben Layne
Barbara #RightMatters
Raquel Lucio Lula da Silva
Sande
Jon Murray
hire urban meyer
Opa'ivu'eke stan
M.Bayne 👻🦴🧠🍁🎃🍂 Boo!
Hoss Tagge, PhD
NancyJohnson
Sean McCabe
Amanda J Wallace
Oscar Vivas
mglservicios365@gmail.com
Matías Carrancio
Shekinah
cajuuu_ctn
Frontier Investor
industriam adiuvat deus
Ila Treat
Anarchist B&B
VOTING IS PATRIOTIC
Resist 11
Jorge FroylanPSYE
Sam Hollingsworth
sumi__
pat micheletti
Oscar Johnson
han 🧝
Bettie Marchiori
Josh Dobbins.
Indivisible
Lily de Vries
Shane A Stockdale
psm
Angel pucciarelli
Deb 🌺
Ethan Stewart
Kathleen Anger
B.A.Resister
DocJay *BLM*D
@centauro1262
nope
Lasso Foolek
vikki manderson
Susan DeSanti
Social Distancing Roga
Roi 🏳️‍🌈
Anna Bellard 🏡😊stay healthy 🍑again!
Marie Antoinette
Ad/SC BR João 8;32 BR
HRH The King of Bohemia, Holy Roman Emperor
Not a Canberra Bubbler 🌲🌍🌲
Robo(not a real cop) Fartlander
the vidya game 🅱️oy
Lila
Christine_Sage

DonnaRocky
Adriana 2 4 days to go
MagentaSky_234
Dillon Lee
No Lockdown
Pastor Mary McQueen
Paulo
eleanor pepper
Le Dédômisé
JacquelineH 🙏
Mairead Ni Rodaigh
zandra
Mari's A Nationalist US❌⭐⭐⭐IT
💜ⒷⒺⓃⒶⓂⒿⓄⓄⓃⒾⒺ i⬜⬜💜
Miri Damas 🌺
ᴮᴱMary Therese⁷⬜ ⬜⬜-⁷
Kim Thomas
justsayOM
LP_LP
Diego Valencia Ⓜ
イマン

THATjamesEtierney🍀🤞
chandler
Alex Ter Neopolitica #UCR
t!j✏⁷
thecommish27
Abdullah
MIPPCI
Dr Joe Pajak CSci NHS Foundation Trust Governor
Breeeeeeeee
Felipe E. González
Evelyn Ebrey
Silent Majority US CA GB IL
LastAmericanGirl US🌊
Fannie Lou Hamer#Black Lives Matter
Timia
Mr Gman
Nikkol Bull
Jim Stuart
Brian Edey
Tari
nota
Nathan Young 🍀❄

(3) What are the ten most popular hashtags?

In [16]:
```python
hashtag_list = []
for t in data:
    for h in t['entities']['hashtags']:
        if h != []:
            hashtag_list.append(h['text'])
```

In [17]:
```python
pprint(hashtag_list)
```

```
['Canada',
 'COVID19',
 'Alberta',
 'Quebec',
 'covid19',
```

```
  'YouShouldBeHere',
  'Cuba',
  'Miami',
  'USA',
  'DemVoice1',
  'ElCovidMata',
  'LGBTQ',
  'Ahora',
  'coronavirus',
  'BorisJohnson',
  'MattHancock',
  'Cineworld',
  'TESTIMONIO']
```

In [18]:
```python
c = Counter(hashtag_list)
```

In [19]:
```python
c.most_common(10)
```

Out[19]:
```
[('Canada', 1),
 ('COVID19', 1),
 ('Alberta', 1),
 ('Quebec', 1),
 ('covid19', 1),
 ('YouShouldBeHere', 1),
 ('Cuba', 1),
 ('Miami', 1),
 ('USA', 1),
 ('DemVoice1', 1)]
```

(4) Who is the most frequently tweeting person about the topic?

In [20]:
```python
name_list = []
for t in data:
    if t != '':
        name_list.append(t['user']['name'])
```

In [21]:
```python
pprint(name_list)
```

```
['J Donovan',
 'jb',
 'Canadian patriot',
 'BarksBeachesBooks_Kate',
 'D',
 'lastfighter',
 'Shadow',
 'Fiti',
 'Mateo',
 'DiMusa',
 'Sharon Schatz',
 'quack',
 'KS',
 'LALaN🏅🏅健康でいてください👢🍄👢🍄',
 'aloha_kids2',
 'Ship_Love',
 'smilealice12',
 'Maxi Buteler',
```

'A Fair Canada',
'Jim E-H',
'Trace',
'ȷαɱαιcα',
'Chy',
'el unico bolivar',
'Char74Trump',
'Dylan-149',
'Morgan Fitzpatrick, MPH, RN, NP',
'weary_ant',
'Kari Blizzard',
'Open Eugene',
'UnsilentDeaf ᴜs',
'Heerak Christian Kim for US Congress (Virginia-8)',
'Renita',
'Zara 🐾',
'Mary-Ann Noyes',
'Windycity',
'Benjamin Breece',
'Donna 🌍👩🏻\u200d⚕️🦊ᴜs',
'ᴜs Average Deplorable ᴜs',
'🆘Biden / Harris 2020🆘',
'eleanor, 🌹',
'jade nacht',
'Kristin Maier',
'S/o Mlk',
'Diogenes of 白左疆',
'Murphy Payeur',
'Sunday World',
'Capitan Manuel',
'Candice Phelps',
'The Reporters',
'7 estrellas',
'cosima van elgin',
'Renton',
'Dan Johno',
'JJ Gills',
'Roci🌻',
'Marilyn',
'Dave Mattingly',
'Tim',
'Ms. Banks 💝',
'LLB',
'Carlos Toraño',
'BabyBoomer1',
'Matt Aronson',
'Oh for crying out loud! (I am totally a bot)',
'Jobsen20',
'Robots in Kashmir',
'Stefania8',
'ᴍᴇssɪ',
'Tina Wilson',
'Emma - AHA+A+L',
'ᴍɛιʀ☆약속¹³♥',
'Feisty Female',
'Donna Barbieri',
'Jamie Kennedy '
'🏴\U000e0067\U000e0062\U000e0073\U000e0063\U000e0074\U000e007f🏴\U000e0067\U000e0062\U000e0073\U000e0063\U000e0074\U000e007fᴜᴍᴜᴍ',
'FREEPREETY',
'DeeDeeTee',
'Andreina',
'Amriana',

'SmilingAssassin27',
'NETO OLIVEIRA 🌻',
'Richard M Garner',
'Swapnil',
'Beef Queef in YoSelf',
'James',
'Jolly BLM Rogers',
'KGarrison',
'Russell 💣',
'KPM',
'Dr. Allee',
'Infectious Diseases',
'Christian Rosu',
'🦅Readers for Pete👟',
'Aii Ann Kei #RakyatJagaRakyat',
'Jamie Mayerfeld',
'Denia',
'Sadmasist 💀💀',
'TiaLolo',
'Víctor Manuel',
'Shari Sanders',
'Ben Layne',
'Barbara #RightMatters',
'Raquel Lucio Lula da Silva',
'Sande',
'Jon Murray',
'hire urban meyer',
'Opa'ivu'eke stan',
'M.Bayne 👻🦴🍥🍁🎃🍂 Boo!',
'Hoss Tagge, PhD',
'NancyJohnson',
'Sean McCabe',
'Amanda J Wallace',
'Oscar Vivas',
'mglservicios365@gmail.com',
'Matías Carrancio',
'Shekinah',
'cajuuu_ctn',
'Frontier Investor',
'industriam adiuvat deus',
'Ila Treat',
'Anarchist B&B',
'VOTING IS PATRIOTIC',
'Resist 11',
'Jorge FroylanPSYE',
'Sam Hollingsworth',
'sumi__',
'pat micheletti',
'Oscar Johnson',
'han 💁',
'Bettie Marchiori',
'Josh Dobbins.',
'Indivisible',
'Lily de Vries',
'Shane A Stockdale',
'psm',
'Angel pucciarelli',
'Deb 🌺',
'Ethan Stewart',
'Kathleen Anger',
'B.A.Resister',
'DocJay *BLM*D',
'@centauro1262',
'nope',

```
        'Lasso Foolek',
        'vikki manderson',
        'Susan DeSanti',
        'Social Distancing Roga',
        'Roi 🏳\u200d🌈',
        'Anna Bellard 🏠😊stay healthy 🍑again!',
        'Marie Antoinette',
        'Ad/SC ʙʀ João 8;32 ʙʀ',
        'HRH The King of Bohemia, Holy Roman Emperor',
        'Not a Canberra Bubbler 🌲🌐🌲',
        'Robo(not a real cop) Fartlander',
        'the vidya game 🅱oy',
        'Lila',
        'Christine_Sage',
        'DonnaRocky',
        'Adriana 2️⃣4️⃣ days to go',
        'MagentaSky_234',
        'Dillon Lee',
        'No Lockdown',
        'Pastor Mary McQueen',
        'Paulo',
        'eleanor pepper',
        'Le Dédômisé',
        'JacquelineH 🙏',
        'Mairead Ni Rodaigh',
        'zandra',
        "Mari's A Nationalist ᴜꜱ❌⭐⭐⭐ɪᴛ",
        '💜Ⓑⓔⓝⓐⓜⓙⓞⓞⓝⓘⓔ 𝑖🇧🇹💜',
        'Miri Damas 🌸',
        'ᴮᴱMary Therese⁷🇧 🇹🇪-⁷',
        'Kim Thomas',
        'justsayOM',
        'LP_LP',
        'Diego Valencia Ⓜ',
        'イマン',
        'THATjamesEtierney☘✌',
        'chandler',
        'Alex Ter Neopolitica #UCR',
        't!j✏⁷',
        'thecommish27',
        'Abdullah',
        'MIPPCI',
        'Dr Joe Pajak CSci NHS Foundation Trust Governor',
        'Breeeeeeeee',
        'Felipe E. González',
        'Evelyn Ebrey',
        'Silent Majority ᴜꜱ ᴄᴀ ɢʙ ɪʟ',
        'LastAmericanGirl ᴜꜱ🌊',
        'Fannie Lou Hamer#Black Lives Matter',
        'Timia',
        'Mr Gman',
        'Nikkol Bull',
        'Jim Stuart',
        'Brian Edey',
        'Tari',
        'nota',
        'Nathan Young 🍀❄']
```

In [22]:
```python
c = Counter(name_list)
```

```
In [23]:    c.most_common(10)
```

```
Out[23]:  [('J Donovan', 1),
           ('jb', 1),
           ('Canadian patriot', 1),
           ('BarksBeachesBooks_Kate', 1),
           ('D', 1),
           ('lastfighter', 1),
           ('Shadow', 1),
           ('Fiti', 1),
           ('Mateo', 1),
           ('DiMusa', 1)]
```

WordClouds

(1) Visualization with WordCloud of 'frankenstein.txt'

```
In [31]:    # Install
            !pip install wordcloud
```

```
In [32]:    from wordcloud import WordCloud
            import matplotlib.pyplot as plt
            %matplotlib inline
```

```
In [33]:    infile = open('frankenstein.txt')
            contents = infile.read()
            infile.close()
```

```
In [34]:    # Generating a WordCloud image
            wordcloud = WordCloud(max_font_size = 80, collocations = False).generate(contents) # Se
```

```
In [35]:    # Display the generated image:
            %matplotlib inline
            plt.figure(figsize = (15, 10))
            plt.imshow(wordcloud)
            plt.axis("off")
            plt.savefig('WordCloud1.pdf')
            plt.show()
```

(2) Create a WordCloud from Tweets

```
In [36]:   import json
           from wordcloud import WordCloud
           import matplotlib.pyplot as plt
           %matplotlib inline

           #(1) Open the JSON file and create a string variable, content
           infile = open('tweet_stream_covid_200.json')
           data = json.load(infile)
           infile.close()

           contents = ''

           for t in data:
               contents += t['text'] + '\n'

           #(2) Generating a WordCloud image
           wordcloud = WordCloud(max_font_size = 80, collocations = False).generate(contents) # Se

           #(3) Display the generated image:
           %matplotlib inline
           plt.figure(figsize = (15, 10))
           plt.imshow(wordcloud)
           plt.axis("off")
           plt.savefig('WordCloud2.pdf')
           plt.show()
```