realTrump_twetter-Copy1

September 24, 2016

In [318]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import pandas as pd
import csv
from datetime import datetime
import re
from twython import Twython
from twython import TwythonStreamer
import pandas as pd
from collections import Counter
from nltk.corpus import stopwords
import string
import nltk
nltk.download("stopwords")
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/sbuciuma/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[318]: True

In [319]:
```
APP_KEY = "LUpqT9BUMzTzdDFVYMls5myJe"
APP_SECRET = "EP31NVKfAg1kWnE3CA0FRRARJMX07irWX9VQUNyUU2pnFPlBeg"
twitter = Twython(APP_KEY, APP_SECRET)
```

In [320]:
```
for i in range(0, 17):
    user_timeline = twitter.get_user_timeline(screen_name="realDonaldTrum
```

In [321]:
```
#lis=user_timeline[0]['id']-1 #tweet id # for most recent tweet
#only query as deep as necessary
#tweetsum= user_timeline[0]['user']['statuses_count']
#cycles=ceil(tweetsum / 200)
#if cycles>16:
    #cycles=16 #API only allows depth of 3200 so no point trying deeper t
#time.sleep(60)
#for i in range(0, cycles): ## iterate through all tweets up to max of 32
    #incremental = twitter.get_user_timeline(screen_name=handle,
```

```
                #count=200, include_retweets=True, max_id=lis)
                #user_timeline.extend(incremental)
                #lis=user_timeline[-1]['id']-1
                #time.sleep(90) ## 90 second rest between api calls. The API allows 1

In [322]: for i, val in enumerate(user_timeline):
                user_timeline[i]['user_screen_name']=user_timeline[i]['user']['screen
                user_timeline[i]['user_followers_count']=user_timeline[i]['user']['fo
                user_timeline[i]['user_id']=user_timeline[i]['user']['id']
                user_timeline[i]['user_created_at']=user_timeline[i]['user']['created
                if 'retweeted_status' in user_timeline[i].keys():
                    user_timeline[i]['rt_count'] = user_timeline[i]['retweeted_status
                    user_timeline[i]['qt_id'] = user_timeline[i]['retweeted_status']['
                    user_timeline[i]['rt_created'] = user_timeline[i]['retweeted_stat
                    user_timeline[i]['rt_user_screenname'] = user_timeline[i]['retwee
                    user_timeline[i]['rt_user_id'] = user_timeline[i]['retweeted_stat
                    user_timeline[i]['rt_user_followers'] = user_timeline[i]['retweet
                    del user_timeline[i]['retweeted_status']
                if 'quoted_status' in user_timeline[i].keys():
                    user_timeline[i]['qt_created'] = user_timeline[i]['quoted_status'
                    user_timeline[i]['qt_id'] = user_timeline[i]['quoted_status']['id
                    user_timeline[i]['qt_text'] = user_timeline[i]['quoted_status']['
                    user_timeline[i]['qt_user_screenname'] = user_timeline[i]['quoted
                    user_timeline[i]['qt_user_id'] = user_timeline[i]['quoted_status'
                    user_timeline[i]['qt_user_followers'] = user_timeline[i]['quoted_
                    del user_timeline[i]['quoted_status']
                if user_timeline[i]['entities']['urls']: #list
                    for j, val in enumerate(user_timeline[i]['entities']['urls']):
                        urlj='url_'+str(j)
                        user_timeline[i][urlj]=user_timeline[i]['entities']['urls'][j
                if user_timeline[i]['entities']['user_mentions']: #list
                    for j, val in enumerate(user_timeline[i]['entities']['user_mentio
                        mentionj='mention_'+str(j)
                        user_timeline[i][mentionj] = user_timeline[i]['entities']['us
                if user_timeline[i]['entities']['hashtags']: #list
                    for j, val in enumerate(user_timeline[i]['entities']['hashtags']):
                        hashtagj='hashtag_'+str(j)
                        user_timeline[i][hashtagj] = user_timeline[i]['entities']['ha
                if user_timeline[i]['coordinates'] is not None:  #NoneType or Dict
                    user_timeline[i]['coord_long'] = user_timeline[i]['coordinates']['
                    user_timeline[i]['coord_lat'] = user_timeline[i]['coordinates']['c
                del user_timeline[i]['coordinates']
                del user_timeline[i]['user']
                del user_timeline[i]['entities']
                if 'place' in user_timeline[i].keys():  #NoneType or Dict
                    del user_timeline[i]['place']
                if 'extended_entities' in user_timeline[i].keys():
                    del user_timeline[i]['extended_entities']
```

```
            if 'geo' in user_timeline[i].keys():
                del user_timeline[i]['geo']
```

```
In [323]: today = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
          print(today)
```

```
2016-09-24 22:44:11
```

```
In [324]: trumpdf = pd.DataFrame(user_timeline)
          #print(dftrump) original dataframe
```

```
In [325]: today = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
          print(today)
          #dftrump['test'] = pd.Series([today for x in range(len(dftrump.index))],
```

```
2016-09-24 22:44:11
```

```
In [326]: #initial creation of dataframe
          #trumpdf = pd.DataFrame(user_timeline)
          #convert created date in datetime format
          trumpdf['created_at'] = pd.to_datetime(trumpdf['created_at'])
```

```
In [327]: trumpdf.head()
```

Out[327]:

	contributors	coord_lat	coord_long	created_at	favorite_count
0	None	NaN	NaN	2016-09-24 23:39:02	9874
1	None	NaN	NaN	2016-09-24 17:08:17	39394
2	None	NaN	NaN	2016-09-24 14:10:37	12799
3	None	NaN	NaN	2016-09-24 12:43:52	18974
4	None	NaN	NaN	2016-09-24 12:35:02	34363

	favorited	hashtag_0	hashtag_1	hashtag_2	hashtag_3	...
0	False	AmericaFirst	NaN	NaN	NaN	...
1	False	NaN	NaN	NaN	NaN	...
2	False	NaN	NaN	NaN	NaN	...
3	False	NaN	NaN	NaN	NaN	...
4	False	NaN	NaN	NaN	NaN	...

	retweeted	source	\
0	False	<a href="http://twitter.com/download/iphone" r...	
1	False	<a href="http://twitter.com/download/iphone" r...	
2	False	<a href="http://twitter.com/download/iphone" r...	
3	False	<a href="http://twitter.com/download/android" ...	
4	False	<a href="http://twitter.com/download/android" ...	

	text	truncated	\
0	Thank you Roanoke, Virginia - this a MOVEMENT ...	True	

```
1  If dopey Mark Cuban of failed Benefactor fame ...     False
2  Will be back in Virginia tonight- for a 6pm ra...      True
3  "@KellyannePolls: Trump is headed for a win, s...     False
4  The @SenTedCruz endorsement was a wonderful su...     False

                                                 url_0  \
0                                http://gop.cm/nv5vmm
1                                                 NaN
2  https://twitter.com/i/web/status/7796844688146...
3                                 http://wpo.st/hQg-2
4                                                 NaN

                                                 url_1  \
0  https://twitter.com/i/web/status/7798275133757...
1                                                 NaN
2                                                 NaN
3                                                 NaN
4                                                 NaN

                  user_created_at user_followers_count   user_id  \
0  Wed Mar 18 13:46:38 +0000 2009             11709381  25073877
1  Wed Mar 18 13:46:38 +0000 2009             11709381  25073877
2  Wed Mar 18 13:46:38 +0000 2009             11709381  25073877
3  Wed Mar 18 13:46:38 +0000 2009             11709381  25073877
4  Wed Mar 18 13:46:38 +0000 2009             11709381  25073877

  user_screen_name
0  realDonaldTrump
1  realDonaldTrump
2  realDonaldTrump
3  realDonaldTrump
4  realDonaldTrump

[5 rows x 44 columns]
```

```
In [328]: list(trumpdf.columns.values)
```

```
Out[328]: ['contributors',
           'coord_lat',
           'coord_long',
           'created_at',
           'favorite_count',
           'favorited',
           'hashtag_0',
           'hashtag_1',
           'hashtag_2',
           'hashtag_3',
           'id',
```

```
 'id_str',
 'in_reply_to_screen_name',
 'in_reply_to_status_id',
 'in_reply_to_status_id_str',
 'in_reply_to_user_id',
 'in_reply_to_user_id_str',
 'is_quote_status',
 'lang',
 'mention_0',
 'mention_1',
 'mention_2',
 'mention_3',
 'mention_4',
 'possibly_sensitive',
 'qt_created',
 'qt_id',
 'qt_text',
 'qt_user_followers',
 'qt_user_id',
 'qt_user_screenname',
 'quoted_status_id',
 'quoted_status_id_str',
 'retweet_count',
 'retweeted',
 'source',
 'text',
 'truncated',
 'url_0',
 'url_1',
 'user_created_at',
 'user_followers_count',
 'user_id',
 'user_screen_name']
```

In [329]:
```
#get structure of dataframe: row & col

trumpdf.shape
```

Out[329]: (189, 44)

In [371]:
```
for i in range(0, 7):
    temp_timeline = twitter.get_user_timeline(screen_name="realDonaldTrum
```

In [372]:
```
for i, val in enumerate(temp_timeline):
    temp_timeline[i]['current_time'] = datetime.now().strftime("%Y-%m-%d
    temp_timeline[i]['user_screen_name']=temp_timeline[i]['user']['screen
    temp_timeline[i]['user_followers_count']=temp_timeline[i]['user']['fo
    if 'retweeted_status' in temp_timeline[i].keys():
        temp_timeline[i]['rt_count'] = temp_timeline[i]['retweeted_status
```

```
                temp_timeline[i]['rt_created'] = temp_timeline[i]['retweeted_stat
                temp_timeline[i]['rt_user_id'] = temp_timeline[i]['retweeted_stat
                temp_timeline[i]['rt_user_followers'] = temp_timeline[i]['retweet
                del temp_timeline[i]['retweeted_status']
```

```
In [373]: df3 = pd.DataFrame(temp_timeline)
          df3['created_at'] = pd.to_datetime(df3['created_at'])
```

```
In [374]: df3.shape
```

```
Out[374]: (189, 31)
```

```
In [375]: df3.head()
```

Out[375]:

	contributors	coordinates	created_at	current_time	\
0	None	None	2016-09-24 23:39:02	2016-09-24 22:52:57	
1	None	None	2016-09-24 17:08:17	2016-09-24 22:52:57	
2	None	None	2016-09-24 14:10:37	2016-09-24 22:52:57	
3	None	None	2016-09-24 12:43:52	2016-09-24 22:52:57	
4	None	None	2016-09-24 12:35:02	2016-09-24 22:52:57	

	entities	extended_entities	\
0	{'hashtags': [{'text': 'AmericaFirst', 'indice...	NaN	
1	{'hashtags': [], 'urls': [], 'symbols': [], 'u...	NaN	
2	{'hashtags': [], 'urls': [{'display_url': 'twi...	NaN	
3	{'hashtags': [], 'urls': [{'display_url': 'wpo...	NaN	
4	{'hashtags': [], 'urls': [], 'symbols': [], 'u...	NaN	

	favorite_count	favorited	geo	id	...	\
0	10082	False	None	779827513375780864	...	
1	39673	False	None	779729180334387200	...	
2	12839	False	None	779684468814667777	...	
3	19042	False	None	779662636770353152	...	
4	34455	False	None	779660413000617985	...	

	quoted_status_id	quoted_status_id_str	retweet_count	retweeted	\
0	NaN	NaN	3277	False	
1	NaN	NaN	15404	False	
2	NaN	NaN	4026	False	
3	NaN	NaN	7603	False	
4	NaN	NaN	10425	False	

	source	\
0	<a href="http://twitter.com/download/iphone" r...	
1	<a href="http://twitter.com/download/iphone" r...	
2	<a href="http://twitter.com/download/iphone" r...	
3	<a href="http://twitter.com/download/android" ...	
4	<a href="http://twitter.com/download/android" ...	

```
                                                      text truncated  \
0  Thank you Roanoke, Virginia - this a MOVEMENT ...      True
1  If dopey Mark Cuban of failed Benefactor fame ...     False
2  Will be back in Virginia tonight- for a 6pm ra...      True
3  "@KellyannePolls: Trump is headed for a win, s...     False
4  The @SenTedCruz endorsement was a wonderful su...     False

                                                      user user_followers_count
0  {'notifications': None, 'created_at': 'Wed Mar...             11709478
1  {'notifications': None, 'created_at': 'Wed Mar...             11709478
2  {'notifications': None, 'created_at': 'Wed Mar...             11709478
3  {'notifications': None, 'created_at': 'Wed Mar...             11709478
4  {'notifications': None, 'created_at': 'Wed Mar...             11709478

  user_screen_name
0  realDonaldTrump
1  realDonaldTrump
2  realDonaldTrump
3  realDonaldTrump
4  realDonaldTrump

[5 rows x 31 columns]
```

```
In [376]: list(df3.columns.values)
```

```
Out[376]: ['contributors',
           'coordinates',
           'created_at',
           'current_time',
           'entities',
           'extended_entities',
           'favorite_count',
           'favorited',
           'geo',
           'id',
           'id_str',
           'in_reply_to_screen_name',
           'in_reply_to_status_id',
           'in_reply_to_status_id_str',
           'in_reply_to_user_id',
           'in_reply_to_user_id_str',
           'is_quote_status',
           'lang',
           'place',
           'possibly_sensitive',
           'quoted_status',
           'quoted_status_id',
           'quoted_status_id_str',
```

```
 'retweet_count',
 'retweeted',
 'source',
 'text',
 'truncated',
 'user',
 'user_followers_count',
 'user_screen_name']
```

In [377]:
```
#create create dataframe with the required fields for analysis.
#trumpdf is the primary/original tweeter request, we create the df1
df1 = trumpdf[['created_at','favorite_count','id', 'lang', 'retweet_count
```

In [378]:
```
df1.shape
```

Out[378]:
```
(189, 8)
```

In [379]:
```
#get the index/column list
list(df1.columns.values)
```

Out[379]:
```
['created_at',
 'favorite_count',
 'id',
 'lang',
 'retweet_count',
 'user_followers_count',
 'user_screen_name',
 'text']
```

In [380]:
```
#create second dataframe witht he required fields for join and create the
#so we create df2

df2 = df3[['current_time','favorite_count','id', 'retweet_count', 'user_f
```

In [381]:
```
list(df2.columns.values)
```

Out[381]:
```
['current_time',
 'favorite_count',
 'id',
 'retweet_count',
 'user_followers_count',
 'user_screen_name']
```

In [382]:
```
df2.shape
```

Out[382]:
```
(189, 6)
```

In [383]:
```
realDonaldTrump = pd.merge(df1, df2, how='inner', on=['id', 'user_screen_
```

In [384]:
```
#create final dataframe with required fields for analysis. using inner jo
#realDonaldTrump = pd.merge(df1, df2, how='inner', on=['key1', 'key2'])
```

```
In [385]: realDonaldTrump.head()
```

Out[385]:

	created_at	favorite_count_x	id	lang
188	2016-09-01 04:58:53	18242	771210697552166913	en
187	2016-09-01 05:06:52	20761	771212706275741696	en
186	2016-09-01 05:07:04	23611	771212757517557760	en
185	2016-09-01 10:31:17	60412	771294347501461504	en
184	2016-09-01 10:40:13	29179	771296597963661312	en

	retweet_count_x	user_followers_count_x	user_screen_name
188	6502	11709381	realDonaldTrump
187	7126	11709381	realDonaldTrump
186	8726	11709381	realDonaldTrump
185	30454	11709381	realDonaldTrump
184	7082	11709381	realDonaldTrump

	text	current_ti
188	There will be no amnesty!\n#MakeAmericaGreatAg...	2016-09-24 22:52:
187	Hillary Clinton doesn't have the strength or t...	2016-09-24 22:52:
186	Under a Trump administration, it's called #Ame...	2016-09-24 22:52:
185	Mexico will pay for the wall!	2016-09-24 22:52:
184	Thank you to @foxandfriends for the great revi...	2016-09-24 22:52:

	favorite_count_y	retweet_count_y	user_followers_count_y
188	18242	6502	11709478
187	20761	7126	11709478
186	23611	8726	11709478
185	60412	30455	11709478
184	29179	7082	11709478

```
In [386]: #d = dict(screen_name = screen_name, Favorites = favorites_count, Retweet

          #HillaryClinton = pd.DataFrame({k : pd.Series(v) for k, v in  list(d.iter
          status_texts = realDonaldTrump[['text']]
          print(status_texts)
```

```
                                                  text
188  There will be no amnesty!\n#MakeAmericaGreatAg...
187  Hillary Clinton doesn't have the strength or t...
186  Under a Trump administration, it's called #Ame...
185                      Mexico will pay for the wall!
184  Thank you to @foxandfriends for the great revi...
183         Poll numbers way up - making big progress!
182  Thank you for having me this morning @American...
181  I am promising you a new legacy for America. W...
180  I will be interviewed by @ericbolling tonight ...
179  Just heard that crazy and very dumb @morningmi...
178  People will be very surprised by our ground ga...
```

```
177  I visited our Trump Tower campaign headquarter...
176  Great new poll Iowa - thank you!\n#MakeAmerica...
175   #AmericaFirst #ImWithYou https://t.co/Gtl7DyQkzt
174                 #ImWithYou https://t.co/I8dHzezmY4
173  Thank you Great Faith Ministries International...
172  I am returning to the Pensacola Bay Center in ...
171  Great visit to Detroit church, fantastic recep...
170  .@CNN is so disgusting in their bias, but they...
169  Wow, the failing @nytimes has not reported pro...
168  Crooked Hillary's V.P. pick said this morning ...
167  "@AnneBellar: @realDonaldTrump @CNN CNN is so ...
166  The Republican Party needs strong and committe...
165  The Great State of Arizona, where I just had a...
164  The polls are close so Crooked Hillary is gett...
163  To the African-American community: The Democra...
162  Lyin' Hillary Clinton told the FBI that she di...
161  "@CherNuna: @realDonaldTrump It defies belief ...
160  "@lblackvelvet: @realDonaldTrump We need to sh...
159  "@ronnieclemmons: @ChrisCJackson @TakouiS @rea...
..                                                ...
29   Thank you Georgia! #AmericaFirst\n#MakeAmerica...
28   Thank you Nevada! #AmericaFirst\n#MakeAmericaG...
27   Do people notice Hillary is copying my airplan...
26   Heading to North Carolina for two big rallies...
25   Hillary Clinton is taking the day off again, s...
24   Thank you High Point, NC! I will fight for eve...
23   Thank you Kenansville, North Carolina! Remembe...
22   It is a MOVEMENT - not a campaign. Leaving the...
21   The situations in Tulsa and Charlotte are trag...
20   Hopefully the violence &amp;amp; unrest in Charlot...
19   Thank you Toledo, Ohio! It is so important for...
18   Great new polls! Thank you Nevada, North Carol...
17                      .@YoungDems4Trump  Thank you!
16   "@ThAllenSBoucher: @DiamondandSilk @realDonald...
15   I will be interviewed from Cleveland, Ohio, on...
14                     Will be on @foxandfriends now.
13   Join me in Roanoke, Virginia on Saturday eveni...
12   This is more than a campaign- it is a movement...
11   Spoke with Governor @PatMcCroryNC of North Car...
10   Tomorrow's the day! Knock on doors and make ca...
9    Hillary Clinton just lost every Republican she...
8    'How Trump Would Stimulate the U.S. Economy'\n...
7    Join me in Roanoke, Virginia tomorrow at the B...
6    Crooked Hillary's bad judgement forced her to ...
5    Today is the day! Knock on doors and make call...
4    The @SenTedCruz endorsement was a wonderful su...
3    "@KellyannePolls: Trump is headed for a win, s...
2    Will be back in Virginia tonight- for a 6pm ra...
```

```
1    If dopey Mark Cuban of failed Benefactor fame ...
0    Thank you Roanoke, Virginia - this a MOVEMENT ...

[189 rows x 1 columns]
```

```
In [387]: #count words in each tweet, we can also lower all the words/characters to
          words_trump = pd.Series(' '.join(realDonaldTrump.text).split())
          punctuation = list(string.punctuation)
          stop = stopwords.words('english') + punctuation + ['rt', 'via']
```

```
In [388]: #implementation of stop words using nltk library for language processing
          terms_stop = [term for term in words_trump if term not in stop]
```

```
In [389]: print(terms_stop)
```

```
['There', 'amnesty!', '#MakeAmericaGreatAgain', '#ImWithYou', 'https://t.co/vVhzSdC
```

```
In [390]: word_counts = {}
          for word in terms_stop:
              if word in word_counts:
                  word_counts[word] += 1
              else:
                  word_counts[word] = 1
```

```
In [391]: print(word_counts)
```

```
{'every': 2, 'https://t.co/4iW06ROvsT': 1, 'https://t.co/bnEE6NX41Z': 1, '@oreillyf
```

```
In [392]: #one more filtering with some custome specified stopwords what are not in
          filterTump = {k: v for k, v in word_counts.items() if v > 1}

          print(filterTump)

          Stopwords = ['A','AND','An','And','As','Be','C','this','they',
          'Can','D','Do','Don','For','Go','He','IS','Is','It','K','M', 'I',
          'O','On','P','Q','R','S','So','T','THE','That','The','Their','There',
          'They','This','U','we','you','words','w','ve','u','the','that','than',
          'her','he','had','for','e','d','co','a','V','TV','was','to','so','she'
          't','such','some','s','re','my','m','ll','is','it','of','as','at','am',
          't', 'or','our','in','do','be','them','they','their','this','were','when'
          'who','with','what','amp','an','and','are','000','your','she','him','his'
          'get','but','would','https','on','realDonaldTrump','have','will','all','h
          'just', 'MakeAmericaGreatAgain','now','out','about','from','by', 'the', '

          for k, v in list(filterTump.items()):
              print(k, v)
              if k in Stopwords:
                  del filterTump[k]
```

{'every': 2, 'bad': 2, 'MAKE': 9, 'Together,': 4, 'Friday': 3, 'morning': 2, 'took'
every 2
bad 2
MAKE 9
Together, 4
Friday 3
morning 2
took 2
last 9
VOTE 2
Thank 41
Action! 2
put 2
https://t.co/3KWOl2ibaW. 2
SO 2
going 5
Iowa 2
#MAGA 11
Americans 2
disgusting 2
working 2
President 4
#AmericaFirst! 2
told 2
Trump 10
Secretary 2
families 2
people 6
noon! 2
including 2
amazing 2
hope 2
#ImWithYou 15
veterans. 2
@realDonaldTrump, 2
bring 2
liked 2
leads 2
officers 2
National 2
#VoteTrump 3
One 2
She 2
knows 4
A 5
doesn't 2
Robert 2
Heading 4

https://t.co/3KWOl2ibaW 5
become 2
Schlafly, 2
Never 4
polls! 2
law 2
great 11
night 2
make 4
They 4
Ohio. 2
things 2
Rep. 2
fantastic 2
6pm! 2
many 5
We 11
you! 10
years. 2
Roanoke, 3
hardly 2
AGAIN! 7
hard 3
Tickets: 2
calls 3
met 2
Ohio 3
badly 2
join 3
AMERICA 10
Hillary's 2
it, 2
failing 6
A.M. 3
Really 2
massive 3
dope! 2
UP 2
.@CNN 3
It 6
tomorrow 3
To 2
honored 2
Day 2
Clive, 2
Great 12
Just 7
Clinton 11

```
Detroit 2
#LaborDay 2
back 5
Video: 2
Carolina 4
MOVEMENT 3
makes 2
military 2
wouldn't 2
stairway 2
Syria 2
American 4
returning 2
@realDonaldTrump 10
get 9
know 4
crowd 3
Enjoy! 10
want 3
Miami 2
International, 2
campaign. 2
poor 2
FBI 2
Wow, 3
soon 2
He 2
thoughts 2
Ohio! 3
veteran 2
getting 2
GREAT 8
attacks 2
night. 4
truly 3
love 2
said 8
plane 2
dopey 2
taken 3
dumb 2
#AmericaFirst 20
ratings 2
deal 2
horrible 2
needs 2
wonderful 6
total 3
```

take 2
results 4
Center 2
See 3
Arizona, 2
interviewed 8
evening 3
fight 3
enjoyed 2
supporters, 2
I 42
ever 3
Florida 4
tonight 5
Florida- 2
Jeff 2
7:00 2
thank 5
-- 2
WIN! 2
Florida, 2
need 3
Commander-in-Chief 3
day! 2
me! 4
Air 2
The 14
future. 2
time 4
like 2
@nytimes 3
together 2
poll 7
big 4
important 2
tremendous 3
rally 5
Americans! 2
Obama 8
would 6
@nytimes, 2
support 2
them. 2
speech 3
https://t.co/TEY6BbCyxt 2
news 2
#MAGA... 2
America 3

years, 2
Will 11
never 7
My 3
Saturday 2
lost 2
wants 3
campaign 2
#TrumpTrain 3
Phyllis 3
failed 4
weak 3
years 2
Mexico 4
endorsement 2
two 2
Forum. 2
today! 2
New 2
major 2
come 2
Under 2
WILL 2
doors 2
job 3
Pensacola, 4
everyone 2
jobs 2
nights 3
Russia 2
me, 3
#MakeAmericaGreatAgain 7
Join 13
Charlotte 3
me. 3
low 2
strong 2
badly. 2
& 20
U.S. 2
new 7
Hillary 23
United 2
much 4
Crooked 8
respected 2
soon. 4
Republican 2

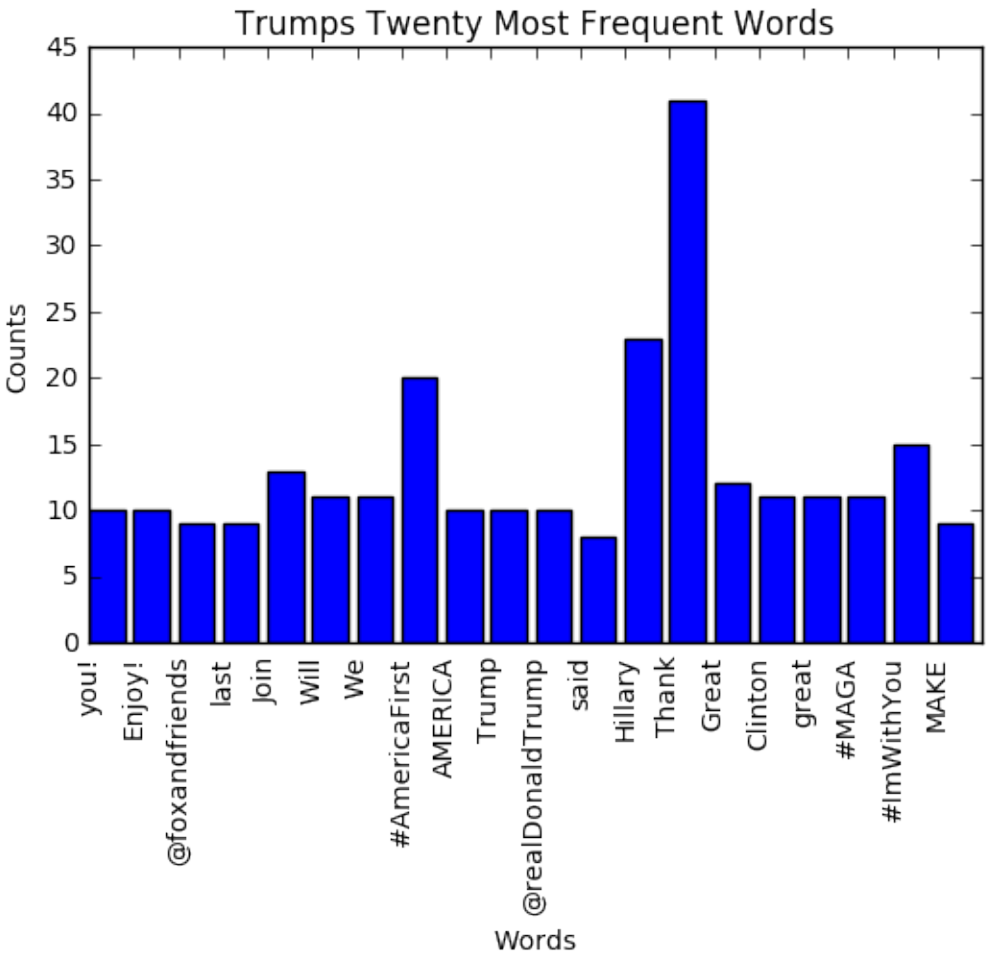
```
@foxandfriends 9
Berglund 2
it's 3
us 7
Governor 2
see 3
@oreillyfactor 4
@seanhannity 2
Gates. 2
don't 2
@GenFlynn 2
polls 3
endorsed 2
https://t.co/L8Ui56dcrJ 2
think 2
reviews 3
@CNN 5
Knock 2
November 3
performance 2
States 2
well 2
now. 2
North 6
prayers 2
country 2
numbers 2
go 3
us: 2
Virginia 4
rally. 2
today 4
voters: 2
vote 3
give 2
```

```
In [393]: toptwentyTrump = dict(Counter(filterTump).most_common(20))
          print(toptwentyTrump)
```

```
{'you!': 10, 'Enjoy!': 10, '@foxandfriends': 9, 'last': 9, 'Join': 13, 'Will': 11,
```

```
In [394]: plt.bar(range(len(toptwentyTrump)), toptwentyTrump.values(), align='edge'
          plt.xticks(range(len(toptwentyTrump)), toptwentyTrump.keys(), rotation=90
          plt.title('Trumps Twenty Most Frequent Words')
          plt.xlabel('Words')
          plt.ylabel('Counts')
```

```
Out[394]: <matplotlib.text.Text at 0x114a520f0>
```




```
In [395]: #counting words
          from collections import Counter
          counts = Counter(word_counts)
          print(counts)
```

```
Counter({'I': 42, 'Thank': 41, 'Hillary': 23, '#AmericaFirst': 20, '&amp;': 20, '#I
```

```
In [396]: a = Counter(filterTump).most_common(20)
          print (a)
```

```
[('Thank', 41), ('Hillary', 23), ('#AmericaFirst', 20), ('#ImWithYou', 15), ('Join'
```

```
In [397]: from prettytable import PrettyTable
          sorted(a)
```

```
          pt = PrettyTable(field_names=['word', 'Frequency_Trump'])
          [ pt.add_row(row) for row in sorted(a, reverse=True)[:20] ]
          pt.max_width['Text'] = 50
          pt.align= 'l'
          print (pt)
```

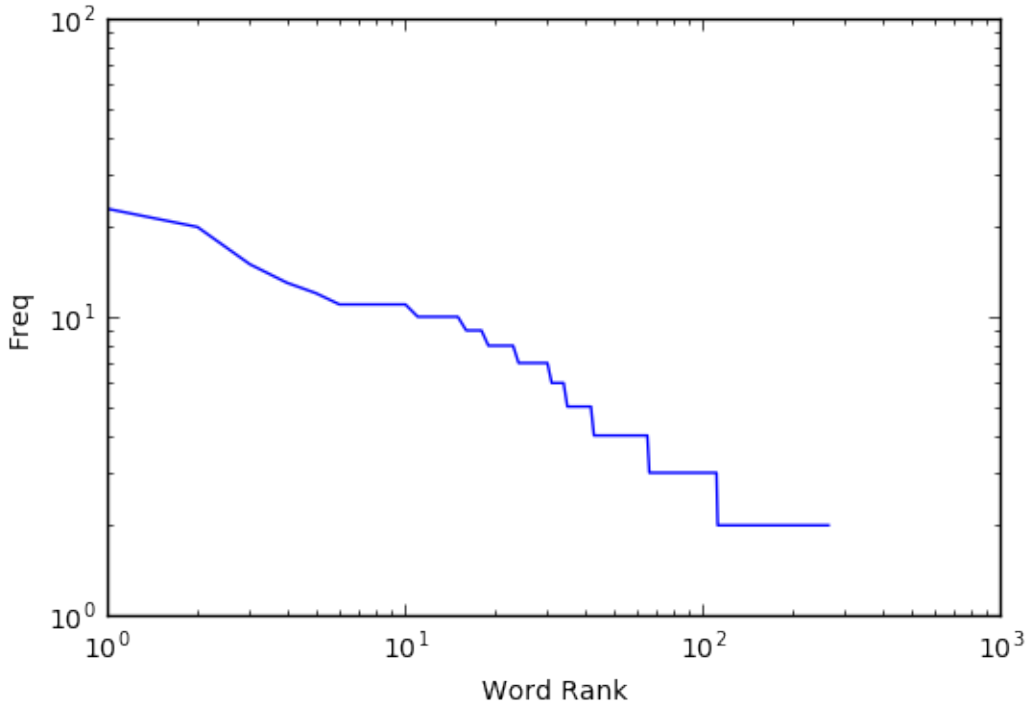
```
+------------------+-----------------+
| word             | Frequency_Trump |
+------------------+-----------------+
| you!             | 10              |
| said             | 8               |
| last             | 9               |
| great            | 11              |
| Will             | 11              |
| We               | 11              |
| Trump            | 10              |
| Thank            | 41              |
| MAKE             | 9               |
| Join             | 13              |
| Hillary          | 23              |
| Great            | 12              |
| Enjoy!           | 10              |
| Clinton          | 11              |
| AMERICA          | 10              |
| @realDonaldTrump | 10              |
| @foxandfriends   | 9               |
| #MAGA            | 11              |
| #ImWithYou       | 15              |
| #AmericaFirst    | 20              |
+------------------+-----------------+
```

In [398]:
```
#Although labels for each word are not provided, x-axis values have been
#Each axis has been adjusted to a logarithmic scale to "squash" the curve
```

In [399]:
```
word_counts = sorted(Counter(filterTump).values(), reverse=True)

plt.loglog(word_counts)
plt.ylabel("Freq")
plt.xlabel("Word Rank")
```

Out[399]: <matplotlib.text.Text at 0x1153afac8>



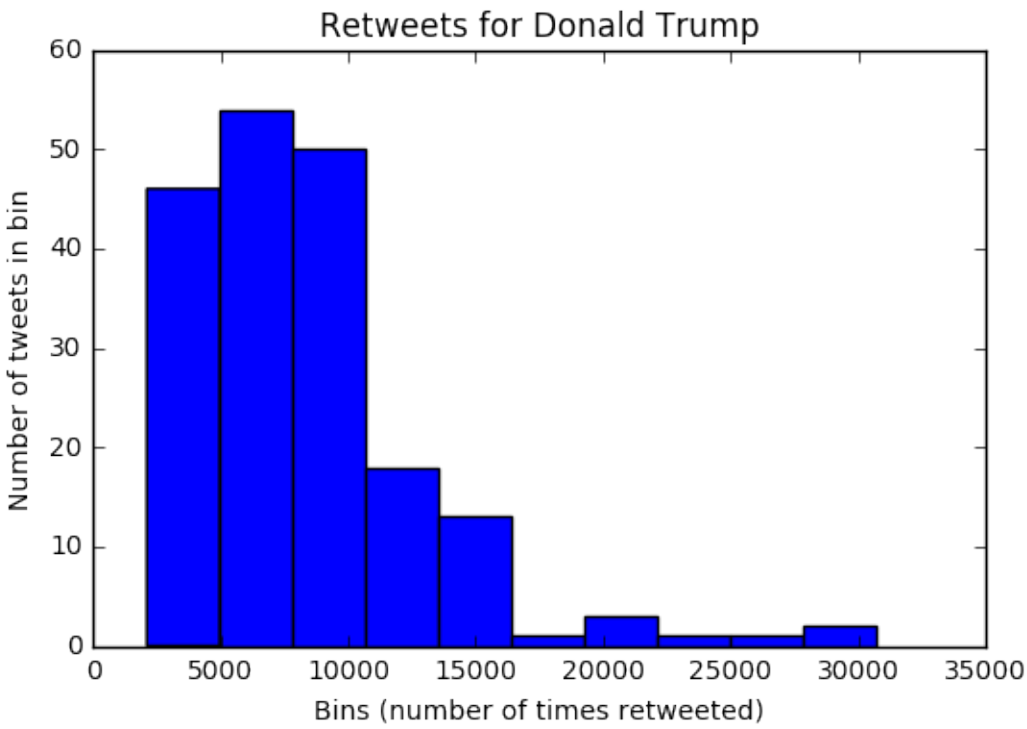

```
In [400]: # Generating a histogram of retweet counts
          #Using underscores while unpacking values in
          # a tuple is idiomatic for discarding them
          retweet = realDonaldTrump['retweet_count_x']

          #print(retweet)
          counts = [count for count in retweet]

          plt.hist(counts)
          plt.title("Retweets for Donald Trump")
          plt.xlabel('Bins (number of times retweeted)')
          plt.ylabel('Number of tweets in bin')

          print (counts)
```

```
[6502, 7126, 8726, 30454, 7082, 10164, 6354, 9717, 3939, 5940, 9395, 7263, 10345, 9
```



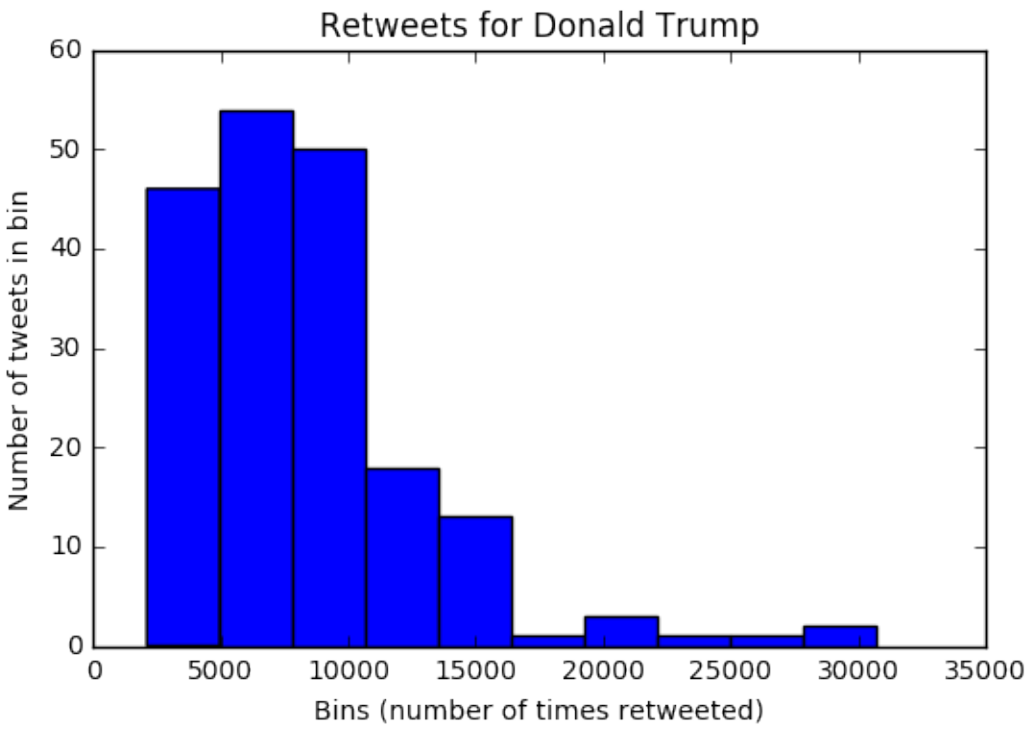

```
In [401]: # Generating a histogram of retweet counts
          #Using underscores while unpacking values in
          # a tuple is idiomatic for discarding them
          retweet = realDonaldTrump['retweet_count_y']

          #print(retweet)
          counts = [count for count in retweet]

          plt.hist(counts)
          plt.title("Retweets for Donald Trump")
          plt.xlabel('Bins (number of times retweeted)')
          plt.ylabel('Number of tweets in bin')

          print (counts)
```

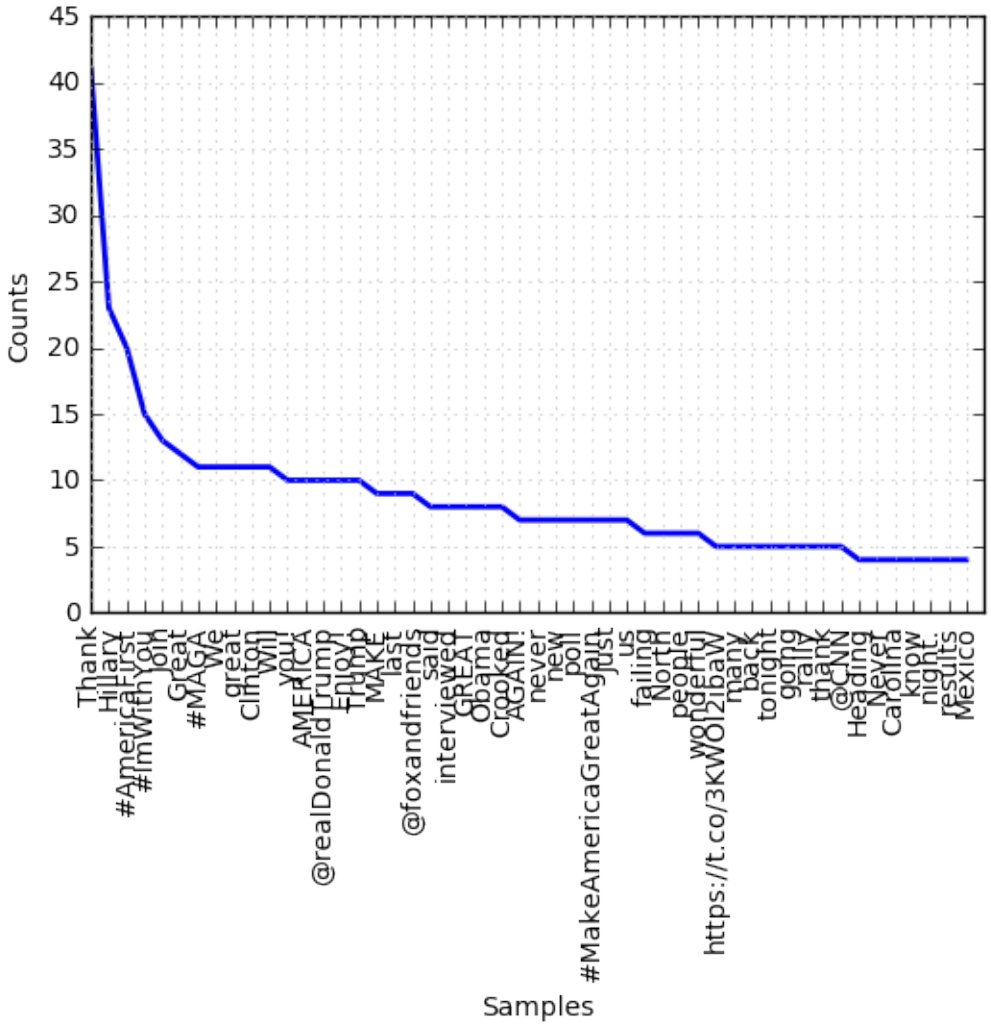
```
[6502, 7126, 8726, 30455, 7082, 10164, 6354, 9717, 3939, 5940, 9395, 7263, 10345, 9
```




```
In [402]: from nltk import FreqDist

          freqdist = nltk.FreqDist(filterTump)

In [403]: freqdist.plot(50)
```




```
In [404]: flyers1m = realDonaldTrump[['created_at','retweet_count_x', 'current_time
          flyers1m.head()
```

Out[404]:

	created_at	retweet_count_x	current_time	retweet_co
188	2016-09-01 04:58:53	6502	2016-09-24 22:52:57	
187	2016-09-01 05:06:52	7126	2016-09-24 22:52:57	
186	2016-09-01 05:07:04	8726	2016-09-24 22:52:57	
185	2016-09-01 10:31:17	30454	2016-09-24 22:52:57	
184	2016-09-01 10:40:13	7082	2016-09-24 22:52:57	

```
In [405]: #create difference in retweets

          realDonaldTrump['retweet_cumulative'] = realDonaldTrump['retweet_count_x'

In [406]: realDonaldTrump.head()
```

Out[406]:

	created_at	favorite_count_x	id	lang
188	2016-09-01 04:58:53	18242	771210697552166913	en
187	2016-09-01 05:06:52	20761	771212706275741696	en
186	2016-09-01 05:07:04	23611	771212757517557760	en
185	2016-09-01 10:31:17	60412	771294347501461504	en
184	2016-09-01 10:40:13	29179	771296597963661312	en

	retweet_count_x	user_followers_count_x	user_screen_name
188	6502	11709381	realDonaldTrump
187	7126	11709381	realDonaldTrump
186	8726	11709381	realDonaldTrump
185	30454	11709381	realDonaldTrump
184	7082	11709381	realDonaldTrump

	text	current_ti
188	There will be no amnesty!\n#MakeAmericaGreatAg...	2016-09-24 22:52:
187	Hillary Clinton doesn't have the strength or t...	2016-09-24 22:52:
186	Under a Trump administration, it's called #Ame...	2016-09-24 22:52:
185	Mexico will pay for the wall!	2016-09-24 22:52:
184	Thank you to @foxandfriends for the great revi...	2016-09-24 22:52:

	favorite_count_y	retweet_count_y	user_followers_count_y
188	18242	6502	11709478
187	20761	7126	11709478
186	23611	8726	11709478
185	60412	30455	11709478
184	29179	7082	11709478

	retweet_cumulative
188	0
187	0
186	0
185	-1
184	0

In [407]:
```
avg = flyers1m.mean()
```

In [408]:
```
print(avg)
```

```
retweet_count_x    8334.560847
retweet_count_y    8336.153439
dtype: float64
```

In [409]:
```
flyers2m = realDonaldTrump[['created_at','retweet_count_x']]
```

In [410]:
```
import vincent
vincent.core.initialize_notebook()
area = vincent.Area(flyers2m)
```

```
        area.colors(brew='Spectral')
        area.display()
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
In [ ]:
```