Hillary_twetter-Copy1

September 24, 2016

In [1]:
```
%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[1]: True

In [2]:
```
APP_KEY = "LUpqT9BUMzTzdDFVYMls5myJe"
APP_SECRET = "EP31NVKfAg1kWnE3CA0FRRARJMX07irWX9VQUNyUU2pnFPlBeg"
twitter = Twython(APP_KEY, APP_SECRET)
```

In [3]:
```
for i in range(0, 17):
    user_timeline = twitter.get_user_timeline(screen_name="HillaryClinton",
```

In [4]:
```
#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 tha
#time.sleep(60)
#for i in range(0, cycles): ## iterate through all tweets up to max of 320
    #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 15

In [5]: for i, val in enumerate(user_timeline):
            user_timeline[i]['user_screen_name']=user_timeline[i]['user']['screen_n
            user_timeline[i]['user_followers_count']=user_timeline[i]['user']['foll
            user_timeline[i]['user_id']=user_timeline[i]['user']['id']
            user_timeline[i]['user_created_at']=user_timeline[i]['user']['created_a
            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']['i
                user_timeline[i]['rt_created'] = user_timeline[i]['retweeted_status
                user_timeline[i]['rt_user_screenname'] = user_timeline[i]['retweete
                user_timeline[i]['rt_user_id'] = user_timeline[i]['retweeted_status
                user_timeline[i]['rt_user_followers'] = user_timeline[i]['retweeted
                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']['te
                user_timeline[i]['qt_user_screenname'] = user_timeline[i]['quoted_s
                user_timeline[i]['qt_user_id'] = user_timeline[i]['quoted_status']|
                user_timeline[i]['qt_user_followers'] = user_timeline[i]['quoted_st
                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_mentions
                    mentionj='mention_'+str(j)
                    user_timeline[i][mentionj] = user_timeline[i]['entities']['user
            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']['hash
            if user_timeline[i]['coordinates'] is not None:  #NoneType or Dict
                user_timeline[i]['coord_long'] = user_timeline[i]['coordinates']['c
                user_timeline[i]['coord_lat'] = user_timeline[i]['coordinates']['co
            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 [10]: today = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
         print(today)
```

```
2016-09-24 21:57:38
```

```
In [11]: HillaryClintondf = pd.DataFrame(user_timeline)
         #print(dftrump) original dataframe
```

```
In [12]: today = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
         print(today)
         #dfHillaryClinton['test'] = pd.Series([today for x in range(len(dftrump.ir
```

```
2016-09-24 21:57:39
```

```
In [13]: #initial creation of dataframe
         #trumpdf = pd.DataFrame(user_timeline)
         #convert created date in datetime format
         HillaryClintondf['created_at'] = pd.to_datetime(HillaryClintondf['created_
```

```
In [14]: HillaryClintondf.head()
```

Out[14]:

	contributors	created_at	favorite_count	favorited	hashtag_0	\
0	None	2016-09-24 23:49:22	4602	False	NaN	
1	None	2016-09-24 22:58:12	3031	False	NaN	
2	None	2016-09-24 21:55:59	5303	False	NaN	
3	None	2016-09-24 21:05:47	4882	False	NaN	
4	None	2016-09-24 19:41:25	5632	False	NaN	

	id	id_str	in_reply_to_screen_name	\
0	779830115643617280	779830115643617280	None	
1	779817238346002432	779817238346002432	None	
2	779801578723217408	779801578723217408	None	
3	779788945953591296	779788945953591296	None	
4	779767715040989184	779767715040989184	None	

	in_reply_to_status_id	in_reply_to_status_id_str	...	\
0	NaN	None	...	
1	NaN	None	...	
2	NaN	None	...	
3	NaN	None	...	
4	NaN	None	...	

	retweeted	source	\
0	False	<a href="https://about.twitter.com/products/tw...	

```
1      False  <a href="https://about.twitter.com/products/tw...
2      False  <a href="https://about.twitter.com/products/tw...
3      False  <a href="https://about.twitter.com/products/tw...
4      False  <a href="https://about.twitter.com/products/tw...

                                                  text truncated  \
0  The new @NMAAHC is an overdue tribute to Afric...     False
1  "Our choice, Hillary Clinton, has a record of ...      True
2  "I thought, this woman--I want to ride with her...     False
3  Hillary's been fighting for women and girls he...     False
4  "A lifetime’s commitment to solving problems i...     False

                                              url_0 url_1  \
0                                               NaN   NaN
1  https://twitter.com/i/web/status/7798172383460...   NaN
2                                               NaN   NaN
3                                http://hrc.io/2d9vSky   NaN
4                               http://nyti.ms/2cVyJ2o   NaN

                   user_created_at  user_followers_count     user_id  \
0  Tue Apr 09 18:04:35 +0000 2013               8951480  1339835893
1  Tue Apr 09 18:04:35 +0000 2013               8951480  1339835893
2  Tue Apr 09 18:04:35 +0000 2013               8951480  1339835893
3  Tue Apr 09 18:04:35 +0000 2013               8951480  1339835893
4  Tue Apr 09 18:04:35 +0000 2013               8951480  1339835893

   user_screen_name
0    HillaryClinton
1    HillaryClinton
2    HillaryClinton
3    HillaryClinton
4    HillaryClinton

[5 rows x 35 columns]
```

```
In [15]: list(HillaryClintondf.columns.values)
```

```
Out[15]: ['contributors',
          'created_at',
          'favorite_count',
          'favorited',
          'hashtag_0',
          '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',
 '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 [16]:
```
#get structure of dataframe: row & col

HillaryClintondf.shape
```

Out[16]:
```
(166, 35)
```

In [17]:
```
for i in range(0, 7):
    temp_timeline = twitter.get_user_timeline(screen_name="HillaryClinton'
```

In [18]:
```
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']['fol
    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_statu
        temp_timeline[i]['rt_user_id'] = temp_timeline[i]['retweeted_statu
        temp_timeline[i]['rt_user_followers'] = temp_timeline[i]['retweete
        del temp_timeline[i]['retweeted_status']
```

In [19]:
```
df3 = pd.DataFrame(temp_timeline)
df3['created_at'] = pd.to_datetime(df3['created_at'])
```

In [20]:
```
df3.shape
```

```
Out[20]: (166, 31)

In [21]: df3.head()

Out[21]:   contributors coordinates          created_at         current_time  \
          0         None        None 2016-09-24 23:49:22  2016-09-24 21:57:53
          1         None        None 2016-09-24 22:58:12  2016-09-24 21:57:53
          2         None        None 2016-09-24 21:55:59  2016-09-24 21:57:53
          3         None        None 2016-09-24 21:05:47  2016-09-24 21:57:53
          4         None        None 2016-09-24 19:41:25  2016-09-24 21:57:53

                                                  entities  \
          0  {'hashtags': [], 'user_mentions': [{'name': 'S...
          1  {'hashtags': [], 'user_mentions': [{'name': 'T...
          2  {'media': [{'source_status_id': 77936203983988...
          3  {'hashtags': [], 'user_mentions': [], 'symbols...
          4  {'hashtags': [], 'user_mentions': [{'name': 'T...

                                         extended_entities  favorite_count  \
          0                                                NaN            4647
          1                                                NaN            3049
          2  {'media': [{'source_status_id': 77936203983988...            5322
          3                                                NaN            4893
          4                                                NaN            5646

            favorited   geo                   id       ...        quoted_status_id  \
          0     False  None  779830115643617280       ...                     NaN
          1     False  None  779817238346002432       ...                     NaN
          2     False  None  779801578723217408       ...                     NaN
          3     False  None  779788945953591296       ...                     NaN
          4     False  None  779767715040989184       ...                     NaN

            quoted_status_id_str  retweet_count retweeted  \
          0                  NaN           1398     False
          1                  NaN           1160     False
          2                  NaN           2238     False
          3                  NaN           1952     False
          4                  NaN           2409     False

                                                   source  \
          0  <a href="https://about.twitter.com/products/tw...
          1  <a href="https://about.twitter.com/products/tw...
          2  <a href="https://about.twitter.com/products/tw...
          3  <a href="https://about.twitter.com/products/tw...
          4  <a href="https://about.twitter.com/products/tw...

                                                text truncated  \
          0  The new @NMAAHC is an overdue tribute to Afric...     False
```

```
1  "Our choice, Hillary Clinton, has a record of ...      True
2  "I thought, this woman--I want to ride with her...     False
3  Hillary's been fighting for women and girls he...     False
4  "A lifetime's commitment to solving problems i...     False

                                                   user user_followers_count
0  {'is_translator': False, 'profile_background_i...               8951497
1  {'is_translator': False, 'profile_background_i...               8951497
2  {'is_translator': False, 'profile_background_i...               8951497
3  {'is_translator': False, 'profile_background_i...               8951497
4  {'is_translator': False, 'profile_background_i...               8951497

  user_screen_name
0   HillaryClinton
1   HillaryClinton
2   HillaryClinton
3   HillaryClinton
4   HillaryClinton

[5 rows x 31 columns]
```

In [22]:
```
list(df3.columns.values)
```

Out[22]:
```
['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 [24]: #create create dataframe with the required fields for analysis.
         #trumpdf is the primary/original tweeter request, we create the df1
         df1 = HillaryClintondf[['created_at','favorite_count','id', 'lang', 'retwe
```

```
In [25]: df1.shape
```

```
Out[25]: (166, 8)
```

```
In [26]: #get the index/column list
         list(df1.columns.values)
```

```
Out[26]: ['created_at',
          'favorite_count',
          'id',
          'lang',
          'retweet_count',
          'user_followers_count',
          'user_screen_name',
          'text']
```

```
In [27]: #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_fo
```

```
In [28]: list(df2.columns.values)
```

```
Out[28]: ['current_time',
          'favorite_count',
          'id',
          'retweet_count',
          'user_followers_count',
          'user_screen_name']
```

```
In [29]: df2.shape
```

```
Out[29]: (166, 6)
```

```
In [30]: realHillaryClinton = pd.merge(df1, df2, how='inner', on=['id', 'user_scree
```

```
In [32]: #create final dataframe with required fields for analysis. using inner joi
         #realHillaryClinton = pd.merge(df1, df2, how='inner', on=['key1', 'key2'])
```

In [33]:
```
realHillaryClinton.head()
```

Out[33]:

	created_at	favorite_count_x	id	lang	\
165	2016-09-16 13:58:03	1920	776782199832113152	en	
164	2016-09-16 13:58:45	1890	776782378375208960	en	
163	2016-09-16 13:59:10	2812	776782482352009216	en	
162	2016-09-16 14:00:39	2383	776782857993916416	en	
161	2016-09-16 14:01:46	1700	776783135447089152	en	

	retweet_count_x	user_followers_count_x	user_screen_name	\
165	991	8951480	HillaryClinton	
164	630	8951480	HillaryClinton	
163	1390	8951480	HillaryClinton	
162	1213	8951480	HillaryClinton	
161	840	8951480	HillaryClinton	

	text	current_tim
165	"We know who Donald is. For five years, he has...	2016-09-24 21:57:5
164	Just yesterday, Trump again refused to say wit...	2016-09-24 21:57:5
163	Barack Obama was born in America, plain and si...	2016-09-24 21:57:5
162	Donald Trump is unfit to be president.\nWe jus...	2016-09-24 21:57:5
161	"Donald Trump looks at a distinguished judge b...	2016-09-24 21:57:5

	favorite_count_y	retweet_count_y	user_followers_count_y
165	1920	991	8951497
164	1890	630	8951497
163	2812	1390	8951497
162	2383	1213	8951497
161	1700	840	8951497

In [34]:
```
#d = dict(screen_name = screen_name, Favorites = favorites_count, Retweets

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

	text
165	"We know who Donald is. For five years, he has...
164	Just yesterday, Trump again refused to say wit...
163	Barack Obama was born in America, plain and si...
162	Donald Trump is unfit to be president.\nWe jus...
161	"Donald Trump looks at a distinguished judge b...
160	"He looks at a Gold Star family and sees them ...
159	Donald Trump looks at women and decides how th...
158	“Let’s reject the cynicism, the bullying, and ...
157	Let’s build a future where love trumps hate. h...
156	What Trump just did is a disgrace.
155	Expressing zero regret for years of pushing a ...

```
154  Leading the birther movement is deplorable. At...
153  When Trump tries to deflect blame for denying ...
152  Again, Trump turns his faults on others. Psych...
151  This shouldn't have to be said: You don’t just...
150  Trump has spent years peddling a racist conspi...
149  The birther lie is what turned Trump from an o...
148  Trump’s birtherism stems from the same innate ...
147  What Trump should do: for once in his life, ow...
146  On National Prisoners of War Remembrance Day, ...
145  .@FLOTUS is on the trail for Hillary. Tune in ...
144  “I am thrilled to be here today to support the...
143  "When it comes to the qualifications we should...
142  "The presidency doesn’t change who you are. It...
141  "Hillary has the resilience it takes to do thi...
140  “We have an opportunity to elect one of the mo...
139  "Elections aren't just about who votes, but wh...
138  "Your decision will determine who sits in the ...
137  “You’ve got to get yourself and everyone you k...
136  “The choice that you make on November 8 will d...
..                                                ...
29   "I think this is the most important election i...
28   "Donald Trump doesn't see people like me, he o...
27   Most people who are bullies grow out of it. \n...
26   The man who could be your next president may b...
25   Very concerned about the outage in Puerto Rico...
24   Donald Trump's ties to Russia may conflict wit...
23   Let's make sure nobody who works full-time in ...
22   "I really regret doing this." --Hillary on Betw...
21   "Vote." --@Lin_Manuel https://t.co/tTgeqxNqYm h...
20   Our favorite moments from Hillary's appearance...
19   "He's 70 years old, and he's still a bully." --...
18   Is this the president we want for our daughter...
17   When Donald Trump speaks about women, our daug...
16   "She gives us hope for the future. She's our g...
15   Charlotte should release police video of the K...
14   Addiction is a disease--not a moral failing. We...
13   "Hillary Clinton would make a sober, smart and...
12   "Donald Trump committed perjury. Or he looked ...
11   "I was just blown away by this 9-year-old girl...
10   Republicans are once again turning their backs...
9    If the young women of this country have Donald...
8    "Hillary Clinton is one of the best prepared c...
7    Donald Trump doesn't see people like Anastasia...
6    RT if you believe that #LoveTrumpsHate\n\nThen...
5    We know all too well who Donald Trump is. Let'...
4    "A lifetime’s commitment to solving problems i...
3    Hillary's been fighting for women and girls he...
2    "I thought, this woman--I want to ride with her...
```

```
1    "Our choice, Hillary Clinton, has a record of ...
0    The new @NMAAHC is an overdue tribute to Afric...

[166 rows x 1 columns]
```

```
In [35]: #count words in each tweet, we can also lower all the words/characters to
         words_HillaryClinton = pd.Series(' '.join(realHillaryClinton.text).split()
         punctuation = list(string.punctuation)
         stop = stopwords.words('english') + punctuation + ['rt', 'via']
```

```
In [36]: #implementation of stop words using nltk library for language processing
         terms_stop = [term for term in words_HillaryClinton if term not in stop]
```

```
In [37]: print(terms_stop)
```

```
['"We', 'know', 'Donald', 'is.', 'For', 'five', 'years,', 'led', 'birther', 'moveme
```

```
{'insults.': 2, 'girl."': 2, 'https://t.co/qa9qO4AFxW': 2, 'disqualify': 2, 'best':
insults. 2
girl." 2
https://t.co/qa9qO4AFxW 2
disqualify 2
best 2
others. 2
-H 4
time 2
let 2
shouldn't 2
himself. 3
future. 3
https://t.co/wDxspmxT9Q 2
history. 2
attacks 2
say 3
It's 3
knocked 2
trumps 3
--@NYTimes 2
tax 2
fighting 2
So 2
afford 2
registered 2
student 3
seen 2
"I 7
need 5
president.” 2
disability. 2
https://t.co/2LRtZB3Ri2 2
“I’m 2
build 6
back 3
inclusive 3
"Hillary 3
progress 2
backs 2
TV 2
watch 2
A 2
president: 2
believes 2
kids 3
can’t 3
like 4
```

economy 4
help 2
people 10
got 7
families 2
matter 2
led 3
without 4
take 5
judge 2
We’ve 2
get 10
would 5
ever 3
fix 2
Obama 2
candidate 2
exactly 3
I 3
better 4
And 2
brave 2
doesn’t 3
We've 2
Tune 2
years 5
Between 2
turning 2
Let's 4
turn 2
country 5
much 5
someone 5
really 2
always 3
doesn't 3
Trump’s 2
said 2
students 2
able 3
America's 2
day 2
black 2
Hillary. 2
peddling 2
made 2
He 2
America, 3

Hillary 15
Flint 2
Trump 41
Barack 3
birther 4
every 4
become 2
needs 2
https://t.co/NG1am536OS 2
sure 5
police 4
ballot. 3
means 2
money 2
this. 4
think 3
five 2
Trump's 2
America 5
stand 2
clear 2
see 5
movement 2
hate. 4
next 4
who’s 2
look 2
Americans 4
“We 2
vote 4
it. 4
way 2
world 2
things 4
are, 2
chance 5
It’s 5
vote." 2
else 2
plan 3
Ferns 2
do. 2
African 3
real 3
--@LATimes 2
too. 2
can't 3
born 4

determine 2
Oval 2
future 3
Our 2
The 5
spent 3
walk 2
still 2
live 3
wrong 3
president 11
November 2
If 4
"Donald 3
used 2
disabilities 2
committed 2
women 6
--@FLOTUS 8
126 2
Let’s 2
Republican 4
unfit 2
first 4
that. 2
one 6
No 3
https://t.co/tTgeqxNqYm 13
In 3
--@NyleDiMarco 2
Clinton 6
far 3
disability 2
Two 2
looks 4
United 2
save 2
another 2
→ 3
other.” 2
know 5
https://t.co/tTgeqy51PU 2
important 2
stop 2
racist 2
--Hillary 15
This 7
deserve 3

presidential 3
we’ve 3
never 3
& 6
good 3
Help 2
intolerable. 2
vote: 2
want 10
campaign 2
You 2
Trump. 2
@POTUS's 3
What 2
We 24
shortchange 2
"You 3
settle 2
going 2
you're 2
part 2
reality 2
release 2
takes 2
greatest 2
are. 2
pragmatic 2
American 7
life 2
country. 2
is. 2
When 5
well 2
Keith 2
voice 2
Donald 26
president. 4
grow 3
love 4
give 3
values 2
using 2
election 4
believe 5
--Roxie, 2
works 2
Hillary's 6
support 3

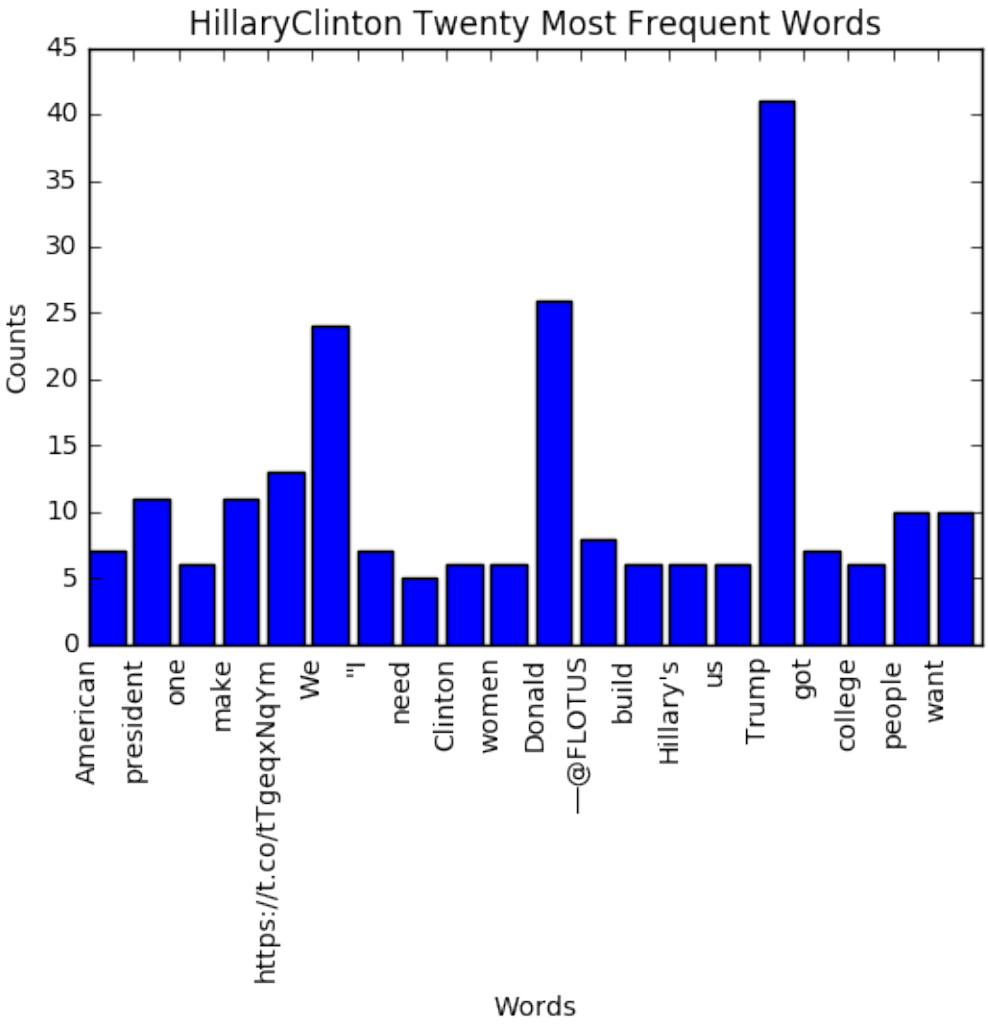
```
make 11
sees 5
Lamont 2
great 5
job 2
https://t.co/s5zyaOBLXB 2
leave 3
people's 2
may 2
share 2
everyone 4
It 5
conspiracy 2
On 3
@POTUS 2
long 3
call 3
gets 5
meeting 2
right 3
change 2
man 2
go 5
it's 3
"We 3
citizenship. 2
something 2
New 2
work 5
regret 2
back. 2
whether 2
I’ve 2
you: 2
"The 3
men 2
us 6
--@FLOTUS: 3
college 6
many 3
terrorists 2
```

```
In [61]: toptwentyHillaryClinton = dict(Counter(filterHillaryClinton).most_common(2
         print(toptwentyHillaryClinton)
```

```
{'American': 7, 'president': 11, 'one': 6, 'make': 11, 'https://t.co/tTgeqxNqYm': 1
```

```
In [62]: plt.bar(range(len(toptwentyHillaryClinton)), toptwentyHillaryClinton.value
         plt.xticks(range(len(toptwentyHillaryClinton)), toptwentyHillaryClinton.ke
         plt.title('HillaryClinton Twenty Most Frequent Words')
         plt.xlabel('Words')
         plt.ylabel('Counts')

Out[62]: <matplotlib.text.Text at 0x115729780>
```




```
In [63]: #counting words
         from collections import Counter
         counts = Counter(word_counts)
         print(counts)

Counter({'Trump': 41, 'Donald': 26, 'We': 24, 'Hillary': 15, '--Hillary': 15, 'http
```

```
In [65]: a = Counter(filterHillaryClinton).most_common(20)
         print (a)

[('Trump', 41), ('Donald', 26), ('We', 24), ('https://t.co/tTgeqxNqYm', 13), ('pres
```

```
In [66]: from prettytable import PrettyTable
         sorted(a)
         pt = PrettyTable(field_names=['word', 'Frequency_HillaryClinton'])
         [ pt.add_row(row) for row in sorted(a, reverse=True)[:20] ]
         pt.max_width['Text'] = 50
         pt.align= 'l'
         print (pt)
```

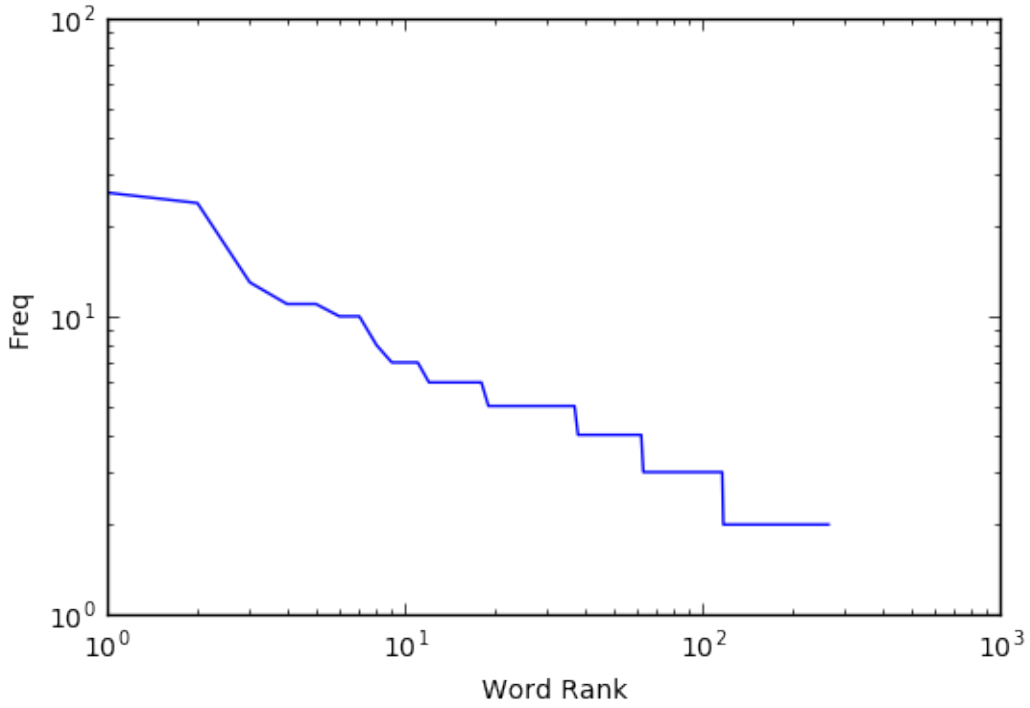
word	Frequency_HillaryClinton
--@FLOTUS	8
women	6
want	10
us	6
president	11
people	10
one	6
need	5
make	11
https://t.co/tTgeqxNqYm	13
got	7
college	6
build	6
We	24
Trump	41
Hillary's	6
Donald	26
Clinton	6
American	7
"I	7

```
In [68]: #Although labels for each word are not provided, x-axis values have been s
         #Each axis has been adjusted to a logarithmic scale to "squash" the curve

In [69]: word_counts = sorted(Counter(filterHillaryClinton).values(), reverse=True)

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

Out[69]: <matplotlib.text.Text at 0x115ded2e8>



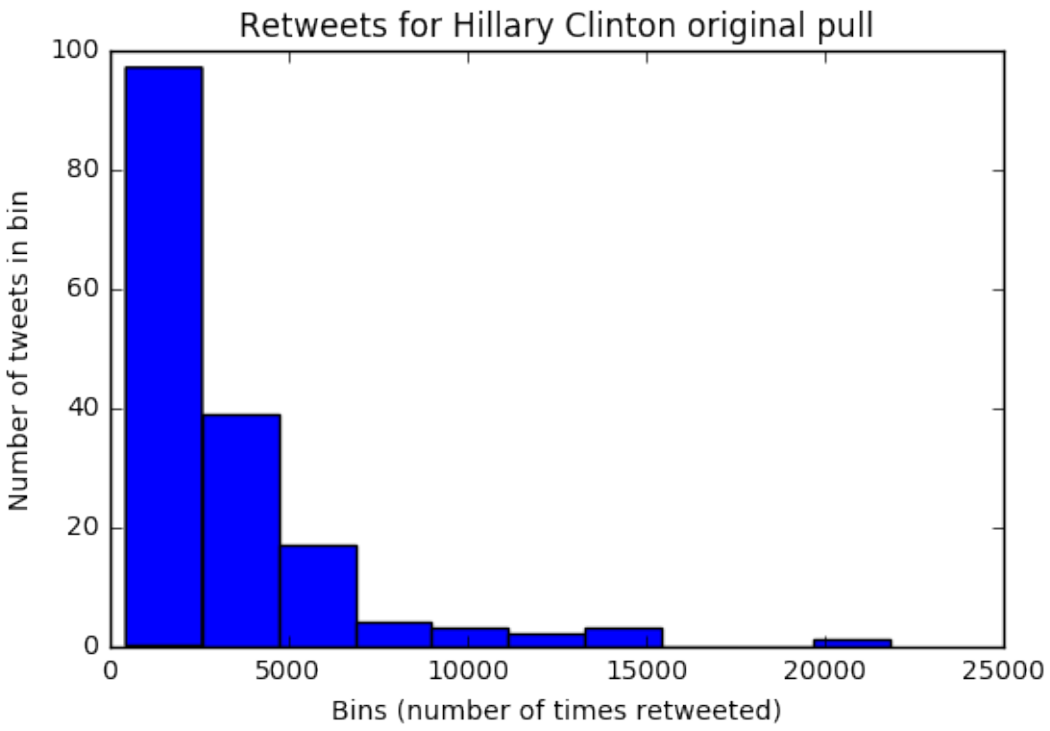
In [77]:
```
# Generating a histogram of retweet counts
#Using underscores while unpacking values in
# a tuple is idiomatic for discarding them
retweet = realHillaryClinton['retweet_count_x']

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

plt.hist(counts)
plt.title("Retweets for Hillary Clinton original pull")
plt.xlabel('Bins (number of times retweeted)')
plt.ylabel('Number of tweets in bin')

print (counts)
```

```
[991, 630, 1390, 1213, 840, 826, 2218, 1190, 999, 9547, 3263, 3316, 3273, 6117, 194
```



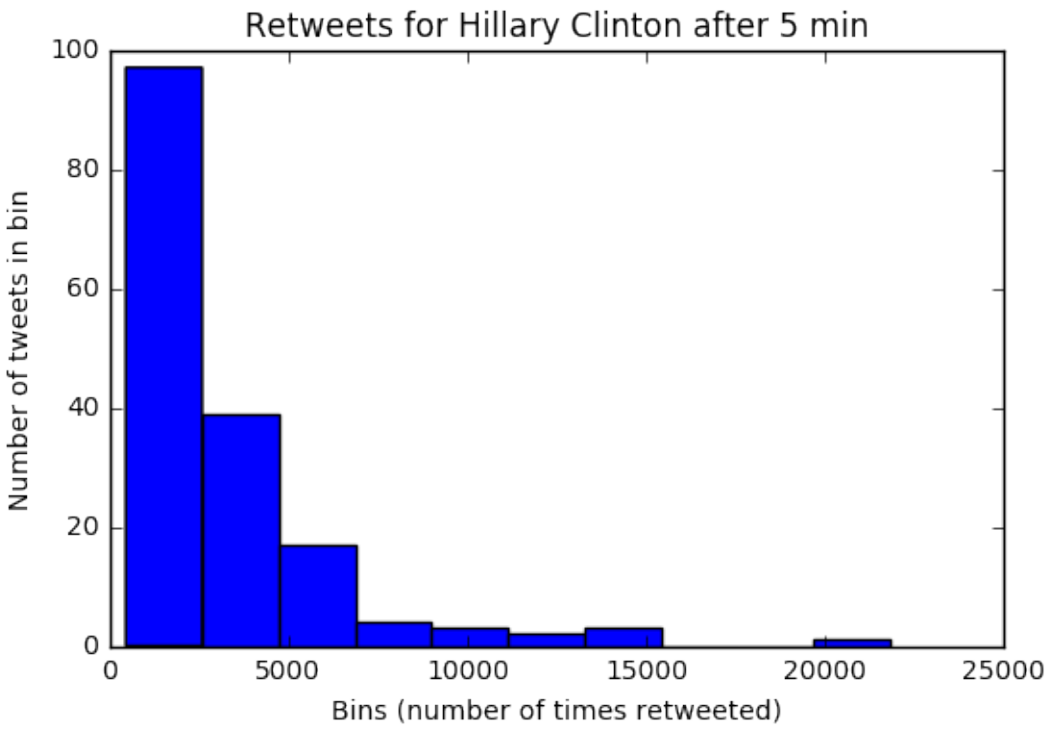

```
In [78]: # Generating a histogram of retweet counts
         #Using underscores while unpacking values in
         # a tuple is idiomatic for discarding them
         retweet = realHillaryClinton['retweet_count_y']

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

         plt.hist(counts)
         plt.title("Retweets for Hillary Clinton after 5 min")
         plt.xlabel('Bins (number of times retweeted)')
         plt.ylabel('Number of tweets in bin')

         print (counts)
```

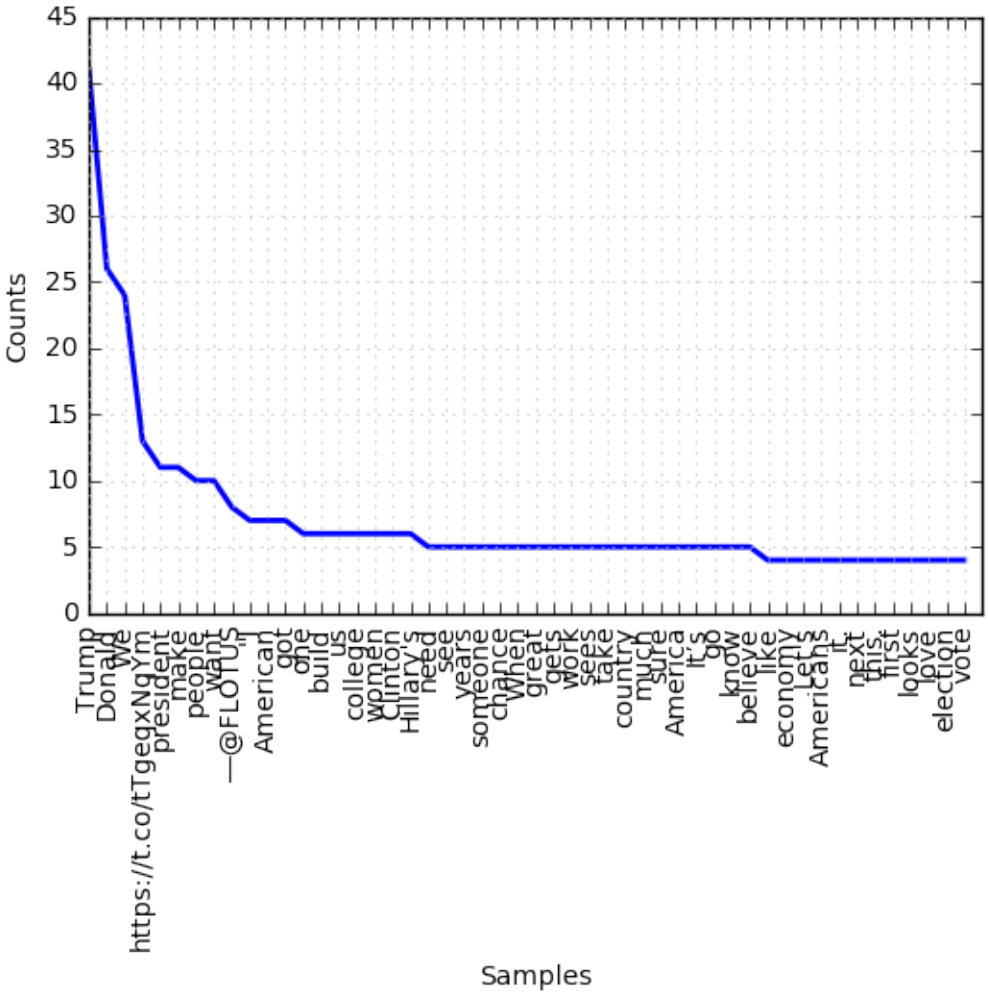
```
[991, 630, 1390, 1213, 840, 826, 2218, 1190, 999, 9547, 3263, 3316, 3273, 6117, 194
```




```
In [79]: from nltk import FreqDist

         freqdist = nltk.FreqDist(filterHillaryClinton)

In [80]: freqdist.plot(50)
```


Counts
Samples
0
5
10
15
20
25
30
35
40
45
Trump
Donald
we
https://t.co/tTgeqxNqYm
president
make
people
want
—@FLOTUS
American
got
one
build
us
college
women
Clinton
Hillary's
need
see
years
someone
chance
When
great
gets
work
sees
take
country
much
sure
America
it's
go
know
believe
like
economy
Let's
Americans
It
next
this
first
looks
love
election
vote

In [ ]: