# Scraping Twitter with Python

Ken Blake, Ph.D.

🕈 Home	Applications on TwitterDemo	✓ Channels		
Environments	¢ jupyter	*	*	lab
Learning	Notebook 6.0.3	console_shortcut 0.1.1	Glueviz 0.15.2	JupyterLab
Community	Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.	Console shortcut creator for Windows (using menuinst)	Multidimensional data visualization across files. Explore relationships within and among related datasets.	An extensible environment for inter and reproducible computing, based Jupyter Notebook and Architectu
	Launch	Install	Install	Install
	¢ IP[y]:	R *	*	A
	Qt Console 4.6.0 PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	RStudio 1.1.456 A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	Spyder 4.0.1 Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features	VS Code 1.42.0 Streamlined code editor with suppo development operations like debug task running and version contro
	Install	Install	Install	Install
Documentation				



Environments	¢	°		lab
Learning	Notebook 6.0.3	console_shortcut 0.1.1	Glueviz 0.15.2 Multidimensional data visualization across	JupyterLab 1.2.5
Community	notebook environment. Edit and run human-readable docs while describing the data analysis.	(using menuinst)	files. Explore relationships within and among related datasets.	and reproducible computing, based of Jupyter Notebook and Architectu
	Launch	Install	Install	Install
	¢ IP[y]:	R *	*	×
	Qt Console	RStudio	Spyder	VS Code
	4.6.0 PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.	1.1.456 A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.	4.0.1 Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features	1.42.0 Streamlined code editor with suppo development operations like debug task running and version contro
	Install	Install	Install	Install
Documentation				







Sign in to Anaconda Cloud
Quit Logout
Upload New - 2
Name   Last Modified File size
14 days ago
14 days ago
a day ago
3 days ago
3 days ago
14 days ago
14 days ago
14 days ago
4 days ago
14 days ago
14 days ago
14 days ago
14 days ago
14 days ago
25 days ago 1.27 kB
~

ð

	C JUpyter TwitterScraper=Ferguson Last Checkpoint: a day ago (unsaved changes)	anda Claud		
	File Edit View Insert Cell Kernel Help			
	H     Run     H     Run     K		Quit	Logout
nents	TwitterScraper Demo		Upload	New 🗸 🙎
		ame 🔸 🛛 Las	t Modified	File size
	A Python program for scraping Twitter content, given search criteria.	14	days ago	
	Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete	14	a days ago	
		:	days ago	
L	In []: N pip install twitterscraper==1.1.0	3	days ago	
nui	In []: 🕅 pip install pandas	14	days ago	
	In []: 🕅 pip install xlsxWriter	14	days ago	
		14	days ago	
	then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.	4	days ago	
		14	days ago	
	In []: M Irom twitterscraper import query_tweets	14	days ago	
	In []: M import datetime as dt import pandas as pd	14	days ago	
		14	days ago	
	Setting parameters for the search. Edit the begin_data, end_date, limit and lang values in the code as desired.	2!	i days ago	1.27 kB
5.	<pre>In []: Degin_date = dt.date(2014,8,9) end_date = dt.date(2014,8,25) </pre>	×		













TwitterScra...





#### TwitterScraper Demo Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. #pip install twitterscraper==1.1.0 In []: Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In a notebook, adding a In [ ]: from twitterscraper import query\_tweets In []: M import datetime as dt import pandas as pd "Markdown" box lets you type Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000notes, instructions and other nonlang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. code content into the program. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read. In []: M df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: #df.shape In []: 🕨 #df.head (20) In []: 🔰 #df.sample (20) In []: 🔰 #df.tail (20) The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data. In []: 🔰 from pandas import ExcelWriter In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete.

- In []:
- #pip install twitterscraper==1.1.0

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

- In [ ]: from twitterscraper import query\_tweets
- In []: M import datetime as dt import pandas as pd

Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired.

#### end date = dt.date(2014, 8, 25)limit = 1000

lang = 'english'

Enter your search term or terms into the code below, between the quote marks after query tweets( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.

In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang) print('Done')

The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.

In []: df = pd.DataFrame(t. dict for t in tweets)

The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.

- In []: #df.shape
- In []: 🕨 #df.head (20)
- In []: 🔰 #df.sample (20)
- In []: #df.tail (20)

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

In []: 🔰 from pandas import ExcelWriter

In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

output.

## "Code" boxes let you edit and run lines of code. They also display

TwitterScraper Demo	
Installing required libraries. These steps are necessary only	/ once per Anaconda environment.
output indicating that the installation is complete.	
n []: 🔰 #pip install twitterscraper==1.1.0	
Importing tools the program needs in order to run. You must run these lines of code every time you use the program line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.	n. To run a
n []: M from twitterscraper import query_tweets	Click on e
[]: N import datetime as dt import pandas as pd	• • ·
Setting parameters for the search. Edit the begin_data, end_date, limit and lang values in the code as desired. []: N begin date = dt.date(2014,8,9)	l its conten
<pre>end_date = dt.date(2014,8,25) limit = 1000 lang = 'english'</pre>	hov thon
Enter your search term or terms into the code below, between the quote marks after query_tweets( and before begindate=begindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u> , for construct complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.	box, then
<pre>]: M tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print('Done')</pre>	the box's
The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dis a type of dataset that Pandas can read.	dataframe
]: M df = pd.DataFrame(tdict for t in tweets)	
The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "(20)" code will show a random sample of 20, if at least 20 are available. The "df.tail (20)" will show the last 20.	t 20 "df.sample
[]: N #df.shape	
[]: N #df.head (20)	
[]: N #df.sample (20)	
[]: 🕅 #df.tail (20)	
The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file or computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' param also customizable. It specifies the name that will be given to the tab in the Excel file where the program will st data.	n your name of meter is tore the
[]: 🕅 from pandas import ExcelWriter	
<pre>[]: Writer = ExcelWriter('Fergusontweets.xlsx') df.to_excel(writer, 'Tweets') writer_excel()</pre>	

o run them, remove the "#" from the code. For each, wait for

## ither type of box to edit t. To run code, click its press Shift/Enter. Or click triangle/bar symbol.

TwitterScraper Demo	
Installing required libraries. These steps are necessary only once poutput indicating that the installation is complete.	per Anaconda environment. To
n [ ]: 🕨 #pip install twitterscraper==1.1.0	
Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.	
<pre>[]: M from twitterscraper import query_tweets []: M import datetime as dt import pandas as pd</pre>	An asteris
Setting parameters for the search. Edit the begin_data, end_date, limit and lang values in the code as desired.	the code is
<pre>[]: &gt;</pre>	actorick di
Enter your search term or terms into the code below, between the quote marks after query_tweets( and before begindate=begindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u> , for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.	
<pre>[]: M tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print('Done')</pre>	finished.
The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.	
<pre>]: M df = pd.DataFrame(tdict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.</pre>	
[]: N #df.shape	
[]: N #df.head (20)	
[]: N #df.sample (20)	
[]: N #df.tail (20)	
The next two code boxes load the ExcelWriter program and save the data in the Fandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.	
[]: 🕅 from pandas import ExcelWriter	
<pre>[]: N writer = ExcelWriter('Fergusontweets.xlsx') df.to_excel(writer, 'Tweets') uriter gaug()</pre>	

o run them, remove the "#" from the code. For each, wait for

## k will appear here while s running. When the sappears, the code has



# These are programs designed to be run by the Python coding



# The code installs them in whatever environment you're using.



writer.save()

## Installation is needed only once per environment. You'll have to delete the # symbols in order to run the code.

A Python program for scraping Twitter content, given search criteria. Installing required branes. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, was tor output indicating that the installation is complete.	iwitters	
Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait bur output indicating that the installation is compute.	A Pythor	program for scraping Twitter content, given search criteria.
<pre>     # fpip install twittercorage==1.1.0     # fpip install pands     # from twitterscraper import query_tweet      [ ]:      # from twitterscraper import query_tweet      [ ]:      # import datetime as dt     import pandas as pd      begindate-begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced/lang-ten, for constructing     complex searches using additional search riteria, hus the code, and wait for a "bone" message to appear.      # tweets = query_tweets("Terguson", begindate-begin_date, endate = end_date, limit = limit, lang = lang)     print("Done")      The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe     is a type of dataset that Pandaa can red.      # df = pd.DataFram(tdict for t in tweets)      The mext line of code are optional. Benow the "#" to run them. The "df.head (20)" code will show the first 20     Fronds in the Pandas dataframe containing the retrieved tweets / just to writy that the program has worked. The 'df.ample     [20]" code will show a random sample of 20, if at least 20 are available. The 'df.till (20)" will show the last 20.      # df.sample (20)      the fact drive. The Even file will be stored in the same directory as this program. You are too pour     compares 'n shord drive. The Even file will be stored in the same directory as this program. You are too pour     for a drive the inserve file will be stored in the same directory as the program. The "youte" parameter is     do accountable. It specifies the name that will be given to the tab in the Evend file where the program vill store the     data. </pre>	Installing requi output indicati	ed libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for g that the installation is complete.
<pre>H #pip install pandes H #pip install xlss#riter Temporting tools the program meds in order to run. You must run these lines of code every time you use the program. To run I # from twitterscraper import query_tweet I : M from twitterscraper import query_tweet I : M import datetime as dt import pandas as pd  begindate-begindate. See Twitter's advanced search tool, https://twitter.com/search-advancedTiang-em, for constructing complex searches using additional search criteria. Bun the code, and wait for a "fone" measage to appear. H tweets = query_tweets("Ferguson", begindate-begin_date, enddate = end_date, limit = limit, lang = lang) print("Dane") The maxt line of code imports the retrieved tweets into a Pandas dataframe. Fandas is a data analysis program. A dataframe is a type of dataset that Pandas can seed. M df = pd.DataFrame(tdictfort in tweets) The maxt line of code are optional. Remove the "#" to run them. The "df head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are available. The "df.taul (20)" will show the last 20. M fef.sample (20) The maxt worked hores lead the Excel/File will be stored in the same directory as this program. You may containt the mase of the fash file (20) The maxt worked hores lead the Excel/File will be stored in the same directory as this program. You may containt the mase of the fash direct. The Excel file will be stored in the same directory as this program. You may containt the mase of the face of file will be stored in the same directory as this program. You may containt the mase of the face of file will be stored in the same directory as this program. You may containt the mase of the face of file will be stored in the same directory as this program. You may containt the mase of the face of file will be stored in the same directory as this program. You may contant the mase of the face of file will be</pre>	]: 🕅 #pip ins	tall twitterscraper==1.1.0
<pre>M frip install xlastriser Temporting tools the program needs in order to run. You must run these lines of code every time you use the program. To run []: M from twitterscraper import query_tweet []: M import datetime as dt import pandas as pd  begindate=Begindate. See Twitter's advanced search tool, <u>https://buitter.com/search-advanced?langTen</u>, for constructing complex searches using additional search criteria. Num the code, and wait for a "Done" measage to appear.  H det = query_tweets("Terguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print("Dose1")  The next line of code imports the retrieved tweets into a Fandas dataframe. Fandas is a data analysis program. A dataframe is a type of dataset that Fandas can read.  M df = pd.DataFrame(tdictfort in tweets)  The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Fandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.ample (20) code will show a random sample of 20, if at least 20 are vaniable. The "df.tail (20)" will show the first 20 records in the Fandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20) code will show a random sample of 20, if at least 20 are vaniable. The "df.tail (20)" will show the last 20.  M faf.tail (20)  The next two code hoxes lead the ExcelWriter program and save the data in the Fandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may outomize the mase of the Excel file by customizing the "Program text will be given to the tab in the Excel file where the program will store the discustomized. The specifies the name that will be given to the tab in the Excel file where the program will store the discustomized. The specifies the name that will be given to the tab in the Excel file where the program will store the discustomized. The spec</pre>	]: 🗎 #pip ins	tall pandas
<pre>Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run []: M from twitterscraper import query_tweet []: M import datetime as dt     import pandas as pd  begindate=Bregindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u>, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. ( ) tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang)  print("Done)  The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read. ( ) df = pd.DataFrame(t,_diot_ for t in tweets)  The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 recordin in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a rundom sample of 20, if at least 20 are available. The "df.head (20)" code will show the last 20. ( ) def.sample (20)  The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your The next wo code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your The next two code boxes load the ExcelWriter program and save the data in the Eanda dataframe to an Excel file on your The next wo code boxes load the ExcelWriter program and save the data in the Eanda dataframe to an Excel file on your Computer's have data file will be stored in the same directory as this program. You any customize the mase of the Excel file by customizing the "remease the two to the tab in the Excel file where the program will store the data.</pre>	]: 🕅 #pip ins	tall xlsxWriter
<pre>[]: ▶ from twitterscraper import query_tweet []: ▶ import datetime as dt import pandas as pd  begindate=begindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=ep</u>, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.</pre>	Importing to	ols the program needs in order to run. You must run these lines of code every time you use the program. To run a
[]: N import datetime as dt import of adaptime as dt import pandas as pd begindate=begindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u> , for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. N tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print("Done") The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read. M aff = pd.DataFrame(tdict_ for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Panda datafarame code in the Panda datafarame code in the "df. anaple (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. M fdf.head (20) M fdf.sample (20) fm enxt two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the "Pempeotweets.kix" name given in the code in the second box. The "Tweets" parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the given in the code in the second box. The "Tweets" parameter is also customizable.	[]:	from twitterscraper import query_tweets
<pre>import pandas as pd begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "bons" message to appear.</pre>	[]:	import datetime as dt
<pre>begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.</pre>		<pre>import pandas as pd</pre>
<pre>begindste=begindste. See Twitter's advanced search tool, https://twitter.com/search-advanced/lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.  W tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print('Done')  The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.  M df = pd.DataFrame(tdict for t in tweets)  The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.  M fdf.sappe  M fdf.head (20)  The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'parent's run in the code in the second box. The 'Tweet's parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.</pre>		
<ul> <li>W tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print('Done')</li> <li>The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.</li> <li>W df = pd.DataFrame(tdict for t in tweets)</li> <li>The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.</li> <li>W fdf.shape</li> <li>M fdf.head (20)</li> <li>M fdf.tail (20)</li> <li>The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by outsomizing the 'remove the same directory as this program. You may customize the name of the Excel file by outsomizing the 'Pengetweets.Xus' name given in the code in the second hox. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.</li> </ul>	complex sear	gindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u> , for constructing ches using additional search criteria. Run the code, and wait for a "Done" message to appear.
The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.	]: N tweets = print('I	<pre>query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) one')</pre>
<pre>M df = pd.DataFrame(tdict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. # #df.shape # #df.head (20) # #df.tail (20) # #df.ta</pre>	The next lin is a type of	e of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe dataset that Pandas can read.
<pre>The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are available. The "df.tail (20)" will show the last 20.  # fdf.shape # fdf.head (20) # fdf.sample (20) # fdf.tail (20) # fd</pre>	]: 🕅 df = pd.	DataFrame(tdict for t in tweets)
<pre># #df.shape #df.head (20) # #df.head (20) # #df.sample (20) # #df.tail (20)  The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.</pre>	The next the records in ( (20)" code (	ee lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 he Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample "ill show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.
<pre># #df.head (20) # #df.sample (20) # #df.tail (20) # #df.t</pre>	[]: 🔰 #df.shar	e
<pre># #df.sample (20) # #df.tail (20) The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.</pre>	]: 🔰 #df.head	(20)
#df.tail (20) The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.	[]: 🔰 #df.samp	le (20)
The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.	]: 🔰 #df.taii	(20)
	The next two computer's I the Excel f: also custom data.	code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your ard drive. The Excel file will be stored in the same directory as this program. You may customize the name of le by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is zable. It specifies the name that will be given to the tab in the Excel file where the program will store the
from pandas import ExcelWriter		

"Import" commands activate parts of the installed programs – sort of like opening Word after having installed Microsoft Office.



writer.save()

# This part of the program allows you to customize what you want to





writer.save()

ify the maximum <sup>F</sup> Tweets you want the capture. The bigger the ne longer the run time.	
ter query_tweets( and before begindate=begindate. Set searches using additional search criteria. Run the c	ee 01
ate = end_date, limit = limit, lang = lang)	



This code limits the search to English-language tweets. Or at least tries to. It doesn't seem to



Type your search term or terms here. This is also the code that, when run, finds and captures the



e advanced search n Twitter, too. The ink will help you build you want.	
ter query_tweets( and before begindate=begindate. Se ex searches using additional search criteria. Run the c	96 01
ate = end_date, limit = limit, lang = lang)	



# These lines import the captured tweets into the Pandas data



## If you'd like to see how many rows and columns of data you got, delete the # and run this line of



# Running this one (minus the #) will show you the first 20 captured



# This one will show you a randomly selected 20 of the tweets you



This one will show you the last 20 tweets you captured. Again, these lines are optional. Remove the # to

#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. In [ ]: #pip install twitterscraper==1.1.0 In []: #pip install pandas In []: M #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In []: M from twitterscraper import query tweets In []: 🔰 import datetime as dt import pandas as pd Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets ( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A datafram is a type of dataset that Pandas can read. In []: df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.samp (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: 🕨 #df.shape from pandas import ExcelWriter In []: In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to excel(writer, 'Tweets') writer.save()

# This last part of the code saves the captured tweets to your computer as an Excel file.

#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. In []: #pip install twitterscraper==1.1.0 In []: #pip install pandas In []: M #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In []: M from twitterscraper import query tweets In []: 🔰 import datetime as dt import pandas as pd Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets ( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A datafram is a type of dataset that Pandas can read. In []: df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.samp (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: 🕨 #df.shape from pandas import ExcelWriter In []: In [ ]: writer = ExcelWriter('Fergusontweets.xlsx') df.to excel(writer, 'Tweets') writer.save()

# Change this text to whatever name you want the saved Excel file to have.

#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. In [ ]: #pip install twitterscraper==1.1.0 In []: #pip install pandas In []: 🕅 #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In []: M from twitterscraper import query tweets In []: 🔰 import datetime as dt import pandas as pd Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets ( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A datafram is a type of dataset that Pandas can read. In []: df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.samp (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: 🕨 #df.shape from pandas import ExcelWriter In []: In [ ]: writer = ExcelWriter('Fergusontweets.xlsx') df.to excel(writer, 'Tweets') writer.save()

## The Excel file will be saved on your computer in the same directory as the program.

#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. In [ ]: #pip install twitterscraper==1.1.0 In []: #pip install pandas In []: 🕅 #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In []: M from twitterscraper import query tweets In []: M import datetime as dt import pandas as pd Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin date, enddate = end date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A datafram is a type of dataset that Pandas can read. In []: df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.samp (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: 🕨 #df.shape from pandas import ExcelWriter In [ ]: In [ ]: Writer = ExcelWriter('Fergusontweets.xlsx') df.to excel(writer, 'Tweets') writer.save()

If there's already an Excel file there with the same name, the original Excel file will be overwritten.

#### A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, we output indicating that the installation is complete.

- In []: #pip install twitterscraper==1.1.0
- In []: 🔰 #pip install pandas
- In []: 🔰 #pip install xlsxWriter

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

- In [ ]: In from twitterscraper import query\_tweets
- In []: M import datetime as dt import pandas as pd

Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired.

#### In []: M begin\_date = dt.date(2014,8,9) end\_date = dt.date(2014,8,25) limit = 1000

lang = 'english'

Enter your search term or terms into the code below, between the quote marks after query\_tweets( and before begindate=begindate. See Twitter's advanced search tool, <a href="https://twitter.com/search-advanced?lang=en">https://twitter.com/search-advanced?lang=en</a>, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.

In []: | tweets = query\_tweets("Ferguson", begindate=begin\_date, enddate = end\_date, limit = limit, lang = lang)
print('Done')

The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.

In []: M df = pd.DataFrame(t.\_\_dict\_\_ for t in tweets)

The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.

- In []: 🕨 #df.shape
- In []: 🔰 #df.head (20)
- In []: 🔰 #df.sample (20)
- In []: 🔰 #df.tail (20)

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

In []: 🕅 from pandas import ExcelWriter

In []: W writer = ExcelWriter('Fergusontweets.xlsx')
df.to\_excel(writer, 'Tweets')
writer.save()

# Here's what running the program looks like.

M Gmail 💪 Google Drive 🖬 Facebook 📑	U IOCAINOST:0000/NOTEDOOKS/DESKTOP/IWITTErScraper/IWITTErScraper%3DFerguson.ipynb	v v	III\ 🗉 🔘 🤠
	Upyter TwitterScraper=Ferguson (autosaved)	ogout	
	File Edit View Insert Cell Kernel Help Trusted Python	n 3 O	
	A A A A A A A A A A A A A A A A A A A		
	TwitterScraper Demo		
	Twitter Scraper Demo		
	A Python program for scraping Twitter content, given search criteria.		
	Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete.		
	In []: N #pip install twitterscraper==1.1.0		
	In []: 🕅 #pip install pandas		
	In []: 🕅 #pip install xlsxWriter		
	Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.		
	In []: M from twitterscraper import query_tweets		
	In []: M import datetime as dt import pandas as pd		
	Setting parameters for the search. Edit the begin_data, end_date, limit and lang values in the code as desired.		
	<pre>In []: M begin_date = dt.date(2014,8,9) end_date = dt.date(2014,8,25) limit = 1000 lang = 'english'</pre>		
	Enter your search term or terms into the code below, between the quote marks after query_tweets( and before begindate=begindate. See Twitter's advanced search tool, <u>https://twitter.com/search-advanced?lang=en</u> , for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.		
	<pre>In []: M tweets = query_tweets("Ferguson", begindate=begin_date, enddate = end_date, limit = limit, lang = lang) print('Done')</pre>		
	The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas		
O Ask me anything		R	へ 宮 印 (× <sup>10:02 AM</sup>

#### A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wa output indicating that the installation is complete.

- In []: #pip install twitterscraper==1.1.0
- In []: 🕅 #pip install pandas
- In []: 🔰 #pip install xlsxWriter

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

- In [ ]: In from twitterscraper import query\_tweets
- In []: M import datetime as dt import pandas as pd

Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired.

#### In []: M begin\_date = dt.date(2014,8,9) end\_date = dt.date(2014,8,25) limit = 1000 lease = leaselish!

lang = 'english'

Enter your search term or terms into the code below, between the quote marks after query\_tweets( and before begindate=begindate. See Twitter's advanced search tool, <a href="https://twitter.com/search-advanced?lang=en">https://twitter.com/search-advanced?lang=en</a>, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.

The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.

In []: M df = pd.DataFrame(t.\_\_dict\_\_ for t in tweets)

The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.

- In []: 🔰 #df.shape
- In []: 🔰 #df.head (20)
- In []: 🔰 #df.sample (20)
- In []: 🔰 #df.tail (20)

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

In []: 🕅 from pandas import ExcelWriter

In []: W writer = ExcelWriter('Fergusontweets.xlsx')
 df.to\_excel(writer, 'Tweets')
 writer.save()

# Time to go look at the Excel file containing the captured tweets.

#### A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, we output indicating that the installation is complete.

- In []: #pip install twitterscraper==1.1.0
- In []: 🕅 #pip install pandas
- In []: 🔰 #pip install xlsxWriter

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

- In [ ]: I from twitterscraper import query\_tweets
- In []: M import datetime as dt import pandas as pd

Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired.

### In []: M begin\_date = dt.date(2014,8,9) end\_date = dt.date(2014,8,25) limit = 1000 lang = 'english'

lang = 'english'

Enter your search term or terms into the code below, between the quote marks after query\_tweets( and before begindate=begindate. See Twitter's advanced search tool, <a href="https://twitter.com/search-advanced?lang=en">https://twitter.com/search-advanced?lang=en</a>, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear.

The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read.

In []: M df = pd.DataFrame(t.\_\_dict\_\_ for t in tweets)

The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.

- In []: 🔰 #df.shape
- In []: 🔰 #df.head (20)
- In []: 🔰 #df.sample (20)
- In []: 🔰 #df.tail (20)

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

In []: 🕅 from pandas import ExcelWriter

In []: W writer = ExcelWriter('Fergusontweets.xlsx')
 df.to\_excel(writer, 'Tweets')
 writer.save()

### I'll also demonstrate a simple framing analysis one could perform.

#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, v output indicating that the installation is complete. In [ ]: #pip install twitterscraper==1.1.0 In []: 🕨 #pip install pandas In []: 🕅 #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In [ ]: ] from twitterscraper import query\_tweets In []: M import datetime as dt import pandas as pd begin date = dt.date(2014, 8, 9)In []: end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code belov search tool, https://twitter.com/search-advanced?la "Done" message to appear. ample In []: In tweets = query tweets("Ferguson", beg print('Done') The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data. In []: 🕅 from pandas import ExcelWriter In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

The captured tweets all mentioned "Ferguson" and appeared during the initial August 2014 wave of unrest in Ferguson, Missouri.

A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, output indicating that the installation is complete.

- In [ ]: #pip install twitterscraper==1.1.0
- In []: #pip install pandas
- In []: 🕅 #pip install xlsxWriter

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

In [ ]: M from twitterscraper import query\_tweets

```
In []: 🔰 import datetime as dt
           import pandas as pd
```

begin date = dt.date(2014, 8, 9) In []: end date = dt.date(2014, 8, 25)limit = 1000lang = 'english'

> Enter your search term or terms into the code belo search tool, https://twitter.com/search-advanced?la "Done" message to appear. ample

> > bed

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

#### In []: 🔰 from pandas import ExcelWriter

In

In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

different frames.

# Some observers called the unrest a "riot." Others called it a "protest." If present, these terms cue very

#### A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, v output indicating that the installation is complete.

- In [ ]: #pip install twitterscraper==1.1.0
- In []: #pip install pandas
- In []: 🕅 #pip install xlsxWriter

Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box.

In [ ]: M from twitterscraper import query\_tweets

```
In []: 🔰 import datetime as dt
           import pandas as pd
```

begin date = dt.date(2014, 8, 9) In []: end date = dt.date(2014, 8, 25)limit = 1000lang = 'english'

> Enter your search term or terms into the code belo search tool, https://twitter.com/search-advanced?la "Done" message to appear. ample

> > bed

The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.

#### In []: 🔰 from pandas import ExcelWriter

In

In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

"protest."

## The analysis looks at which were more common: tweets mentioning "riot," or tweets mentioning



# mentioned neither, but that "protest" tweets significantly outnumbered "riot" tweets.



## A caveat: I did not take the time to find and eliminate duplicate tweets. I have done so in the past, though. The results were similar.

A Python program for scraping Twitter content, given search criteria.

Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wa output indicating that the installation is complete.



Of course, a closer qualitative examination of the tweets would be both possible and wise.

In []: 🕅 from pandas import ExcelWriter

In []: W writer = ExcelWriter('Fergusontweets.xlsx')
 df.to\_excel(writer, 'Tweets')
 writer.save()

C Desktop/TwitterScraper/	K 🧧 TwitterScraper=Ferguson - Jup X +		- a ×
(←) → ℃ @	0 Iocalhost:8888/notebooks/Desktop/TwitterScraper/TwitterScraper%3DFerguson.ipynb	⊠ ☆	III\ 🗊 🔹 👲 😑
M Gmail 💧 Google Drive 📑 Fa	cebook 🚅 Digital Measures 🔞 D2L 🧦 PipeLine 🌴 MTMail 🙅 Associated Press Style 🛅 Data 1 Google Calendar 🔤 GDELT monitor 🌀 SSC 🌴 Power Bl Dashboards J 🜐 Easy Access to NIBRS		
	Cjupyter TwitterScraper=Ferguson (autosaved)	Logout	
	File Edit View Insert Cell Kernel Help	rthon 3 O	
	H     Run     H     Run     C     H     Code     C     m     m     C     m     m     C     m		
		^	^
	/TheFinalWordPod 16 TheFinalWordPod TheFinalWordPodcast 498981981221158912 /status 2014-08-11 Interview Ever 0 0 0 <sup>Cl.</sup> /498981981221158912 #RobinWil	as: js	
	17 KierstPowers Kiersten Powers 498981978239012864 /KierstPowers/status 2014-08-11 LZ: How many unarmed people must die? 0 1 1 Cl. ///www.cl. /498981978239012864 23:59:08 http://u	as: js	
	18 FlowerTea75 FlowerTea Nguyen 498981972572516352 /FlowerTea75/status 2014-08-11 The NAACP meeting /498981972572516352 /498981972572516352 23:59:07 tonight in STL is not what 0 0 1 cl. w	as: js	
	/KingCanaxbi318 19 KingCanaxbi318 �orowelie �orowelie oorowelie oorowe oorowelie oorowelie oorow	as: js	
	<	>	
	In []: <b>M</b> #df.sample (20)		
	In []: 🕅 #df.tail (20)		
	The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Exce file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program w store the data.	el I the Vill	
	In [11]: 🕅 from pandas import ExcelWriter		
	<pre>In [12]: Writer = ExcelWriter('Fergusontweets.xlsx') df.to_excel(writer, 'Tweets') writer.save()</pre>		
	In []: М		



juson Best erview Ever 0 0 0 cla RobinWil	s: js
y unarmed must die? 0 1 1 cla: http://u	s: js
P meeting is not what 0 0 1 cla w	s: js
e big news tions cover 0 0 0 clas Ferguson?	s: js
3	×







#### TwitterScraper Demo A Python program for scraping Twitter content, given search criteria. Installing required libraries. These steps are necessary only once per Anaconda environment. To run them, remove the "#" from the code. For each, wait for output indicating that the installation is complete. In []: #pip install twitterscraper==1.1.0 In []: #pip install pandas In []: M #pip install xlsxWriter Importing tools the program needs in order to run. You must run these lines of code every time you use the program. To run a line of code, click on its box, then press "Shift / Enter." Or, click the triangle/bar symbol next to the box. In []: M from twitterscraper import query tweets In []: M import datetime as dt import pandas as pd Setting parameters for the search. Edit the begin\_data, end\_date, limit and lang values in the code as desired. end date = dt.date(2014, 8, 25)limit = 1000lang = 'english' Enter your search term or terms into the code below, between the quote marks after query tweets ( and before begindate=begindate. See Twitter's advanced search tool, https://twitter.com/search-advanced?lang=en, for constructing complex searches using additional search criteria. Run the code, and wait for a "Done" message to appear. In []: 🔰 tweets = query tweets ("Ferguson", begindate=begin date, enddate = end date, limit = limit, lang = lang) print('Done') The next line of code imports the retrieved tweets into a Pandas dataframe. Pandas is a data analysis program. A dataframe is a type of dataset that Pandas can read. In []: df = pd.DataFrame(t. dict for t in tweets) The next three lines of code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 records in the Pandas dataframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample (20)" code will show a random sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20. In []: #df.shape In []: 🕨 #df.head (20) In []: 🔰 #df.sample (20) In []: #df.tail (20) The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data. In []: 🕨 from pandas import ExcelWriter In []: Writer = ExcelWriter('Fergusontweets.xlsx') df.to\_excel(writer, 'Tweets') writer.save()

## Here's how to save your program and exit both Jupyter Notebook and Anaconda gracefully.

C Desktop/TwitterScraper/ X	TwitterScraper=Ferguson - Jup; × +		- 0 ×
(←) → C' @	0 localhost:8888/notebooks/Desktop/TwitterScraper/TwitterScraper%3DFerguson.ipynb	⊠ ☆	III\ 🗉 🔹 💠 😑
M Gmail 💧 Google Drive 📑 Face	ook 👖 Digital Measures 🔞 D2L 🧦 PipeLine 🌴 MTMail 🙅 Associated Press Style 🚞 Data 1 Google Calendar 🔤 GDELT monitor 🌀 SSC 🧦 Power BI Dashboards J 🔀 Easy Access to NIBRS		
	Cjupyter TwitterScraper=Ferguson (autosaved)	ogout	
	File Edit View Insert Cell Kernel Help Trusted Pythol	on 3 O	
	E + ≫ 20 E ↑ ↓ NRun E C > Code ∨ E		
		^	^
	/TheFinalWordPod 16 TheFinalWordPod TheFinalWordPodcast 498981981221158912 /status /status 2014-08-11 Interview Ever 0 0 0 clas: /498981981221158912 3:59:09 #RobinWil		
	17 KierstPowers Kiersten Powers 498981978239012864 /KierstPowers/status 2014-08-11 LZ: How many unarmed /498981978239012864 23:59:08 people must die? 0 1 1 js http://u		
	18 FlowerTea75 FlowerTea Nguyen 498981972572516352 /FlowerTea75/status 2014-08-11 The NAACP meeting /FlowerTea75/status 2014-08-11 tonight in STL is not what 0 0 1 clas: /498981972572516352 23:59:07 w		
	/KingCanaxbi318 19 KingCanaxbi318 � Towelie � 498981961172795393 /status 2014-08-11 stations cover 0 0 0 clas: /498981961172795393 23:59:04 Ferguson? 5 5	•	
	< >	4	
	In []: N #df.sample (20)		
	In []: 🕅 #df.tail (20)		
	The next two code boxes load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your computer's hard drive. The Excel file will be stored in the same directory as this program. You may customize the name of the Excel file by customizing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is also customizable. It specifies the name that will be given to the tab in the Excel file where the program will store the data.	e	
	In [11]: 🔰 from pandas import ExcelWriter		
	<pre>In [12]: W writer = ExcelWriter('Fergusontweets.xlsx') df.to_excel(writer, 'Tweets') writer.save()</pre>		
	In []: M		



			_		^
Ferguson Best Interview Ever #RobinWil	0	0	0	clas: js	
many unarmed ople must die? http://u	0	1	1	clas: js	
VAACP meeting STL is not what W	0	0	1	clas: js	
of the big news stations cover Ferguson?	0	0	0	clas: js	~
				>	







File Edit View	Insert Cell Kernel Help	The v
New Notebook   Open	↑ ↓ Run ■ C → Markdown ✓ □	
Make a Copy Save as Rename	erScraper Demo	new f
Save and Ch	en en	"Save
Revert to Che		
Print Preview	TwitterScrape/	
Download as	Cancel Save	
Trusted Notebook	o install pandas	
Close and Halt	o install xlsxWriter	
Importin	g tools the program needs in order to run. You must run these lines of code every	
The next three lines of records in the Pandas da (20)" code will show a r	code are optional. Remove the "#" to run them. The "df.head (20)" code will show the first 20 taframe containing the retrieved tweets, just to verify that the program has worked. The "df.sample andom sample of 20, if at least 20 are avaiable. The "df.tail (20)" will show the last 20.	
[]: <b>M</b> #df.shape		
[]: 🔰 #df.head (20)		
[]: ▶ #df.sample (20)		
<pre>[]: M #df.tail (20) The next two code boxes computer's hard drive. T the Excel file by custom also customizable. It sp data.</pre>	load the ExcelWriter program and save the data in the Pandas dataframe to an Excel file on your he Excel file will be stored in the same directory as this program. You may customize the name of izing the 'Pompeotweets.xlsx' name given in the code in the second box. The 'Tweets' parameter is ecifies the name that will be given to the tab in the Excel file where the program will store the	
[]: 🕅 from pandas import E	xcelWriter	

## deo does not show how e your notebook as a ile. But it's the usual as ..." procedure.

# Scraping Twitter with Python

Ken Blake, Ph.D.