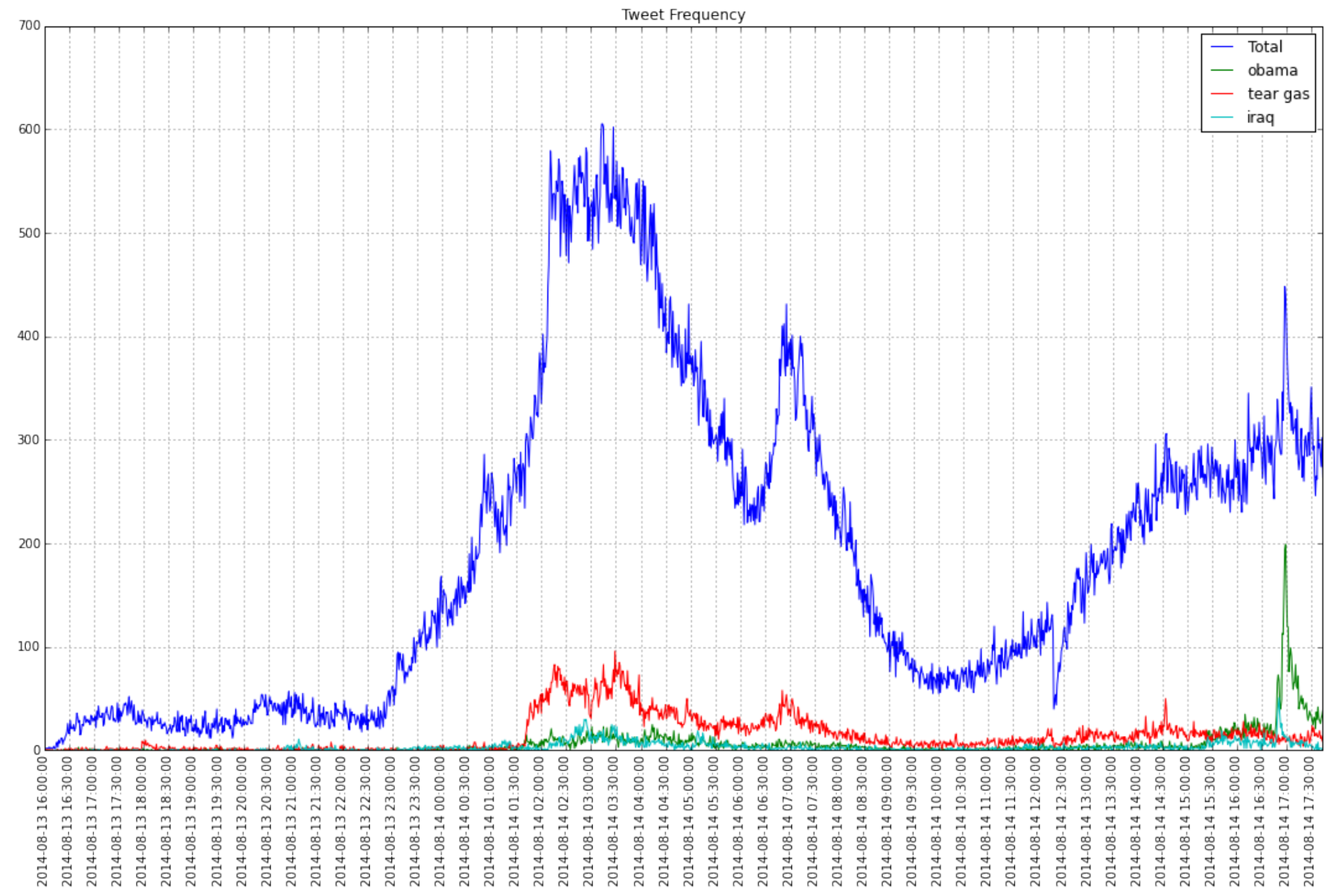
INST728E - MODULE 5

FREQUENCY AND TIME SERIES ANALYSIS

CODY BUNTAIN
@CODYBUNTAIN
CBUNTAIN@CS.UMD.EDU

MODULE 5
FREQUENCY AND TIME SERIES ANALYSIS

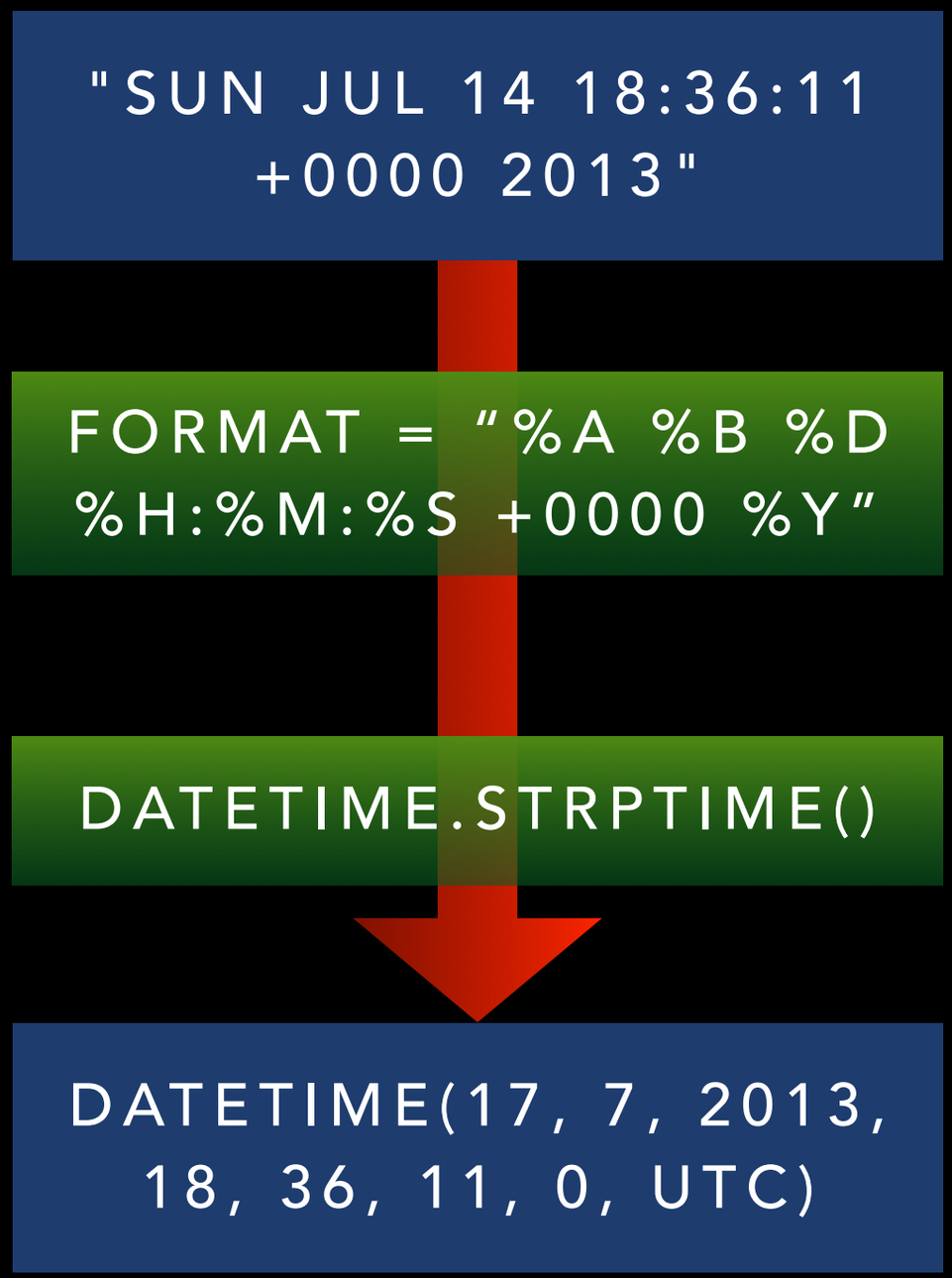



```
Top Twenty Hashtags:
        #ferguson 209701
        #mikebrown 17824
        #mediablackout 5322
        #gaza 4497
        #michaelbrown 2541
        #dontshoot 1968
        #anonymous 1836
        #stl 1607
        #palestine 1542
        #prayforferguson 1525
        #justiceformikebrown 1322
        #opferguson 1160
        #myawhite 995
        #usa 956
        #policestate 906
        #fergusonshooting 875
        #tcot 805
        #inners 773
        #iraq 736
        #fergusonriot 656
```

```
en 282138
es 3759
und 1882
de 1133
tr 795
fr 623
et 476
sk 463
tl 330
in 306
ar 282
it 241
```

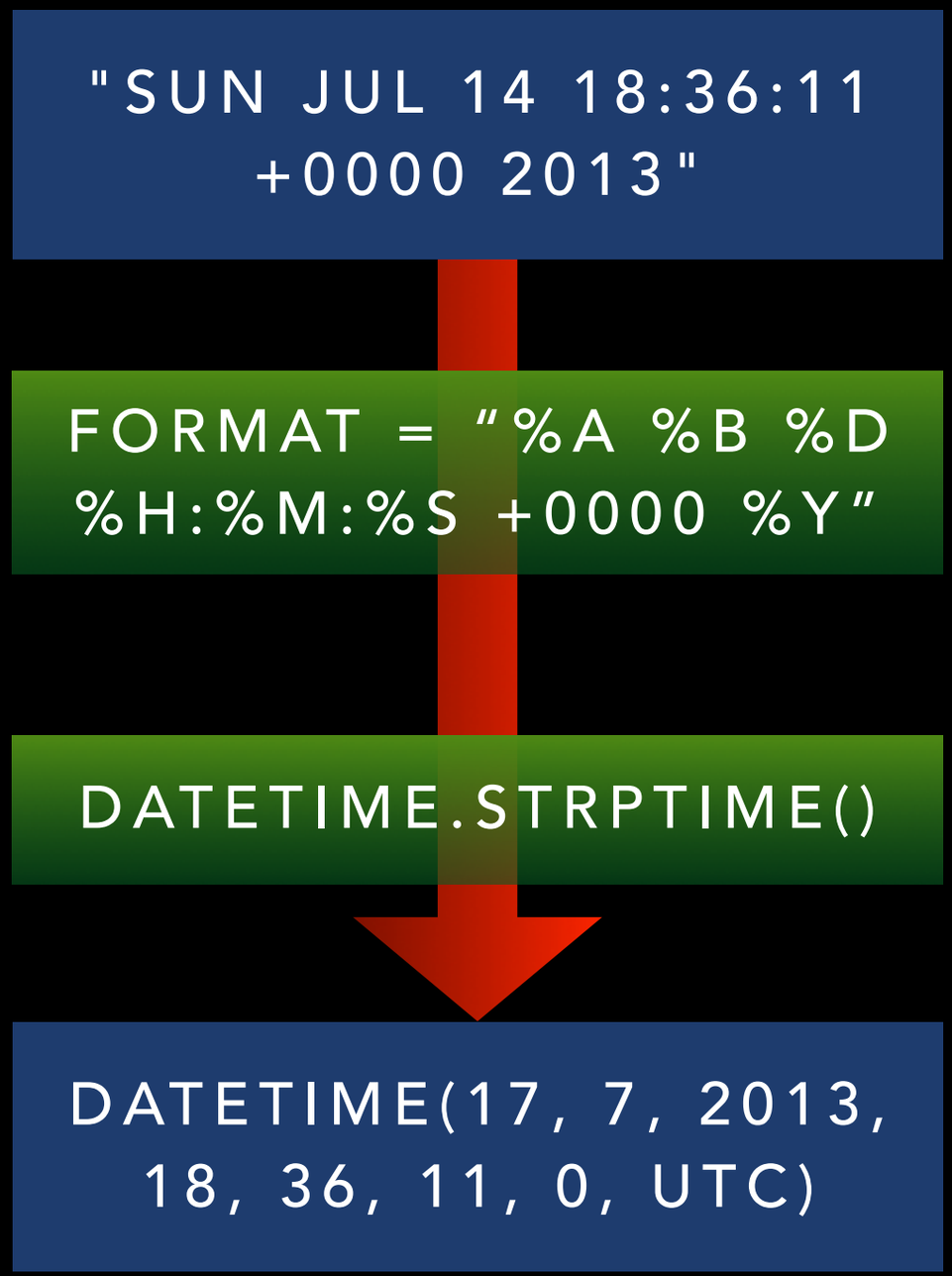
FIRST STEP

- When was a social media message posted?
- Tweets have a "created_at" field
  - e.g., "Sun Jul 14 18:36:11 +0000 2013"
- Want to convert this string to a date and time



FIRST STEP

- Now we can study the number of messages per minute, hour, day, week, etc.



MODULE 5 HOMEWORK

FILL OUT AND SUBMIT MODULE NOTEBOOK

- Using the target event you selected:
    - Choose a few keywords or phrases to describe this event
    - Estimate the time of the disaster using frequency analysis in Twitter
    - Identify the 5 most retweeted messages that were posted ***after*** your selected event

JUPYTER NOTEBOOK

EXAMPLE