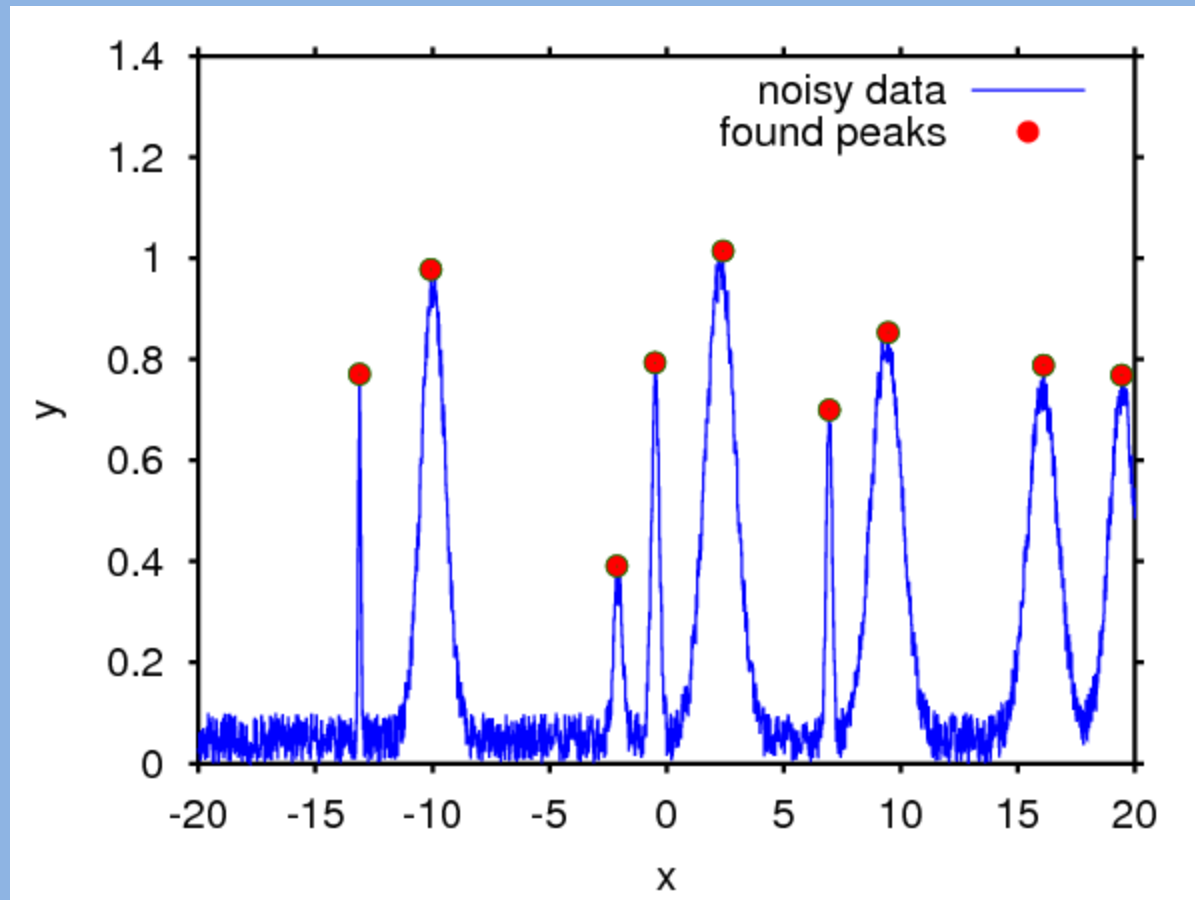


Time Series Use Case

Detecting Peaks...

**Given a long time series,
how do you detect
the top 10 peaks?**



More interesting version of the same question

Given the last 3 years' Tweets per second, how do you detect the top 10 instances when the activity peaked?



YEAR IN REVIEW

Tweets per second

[Home](#)

[The year in stories](#)

[Who joined?](#)

[Hot topics](#)

[Tweets per second](#)

6,939 [JAN 1](#) New Years

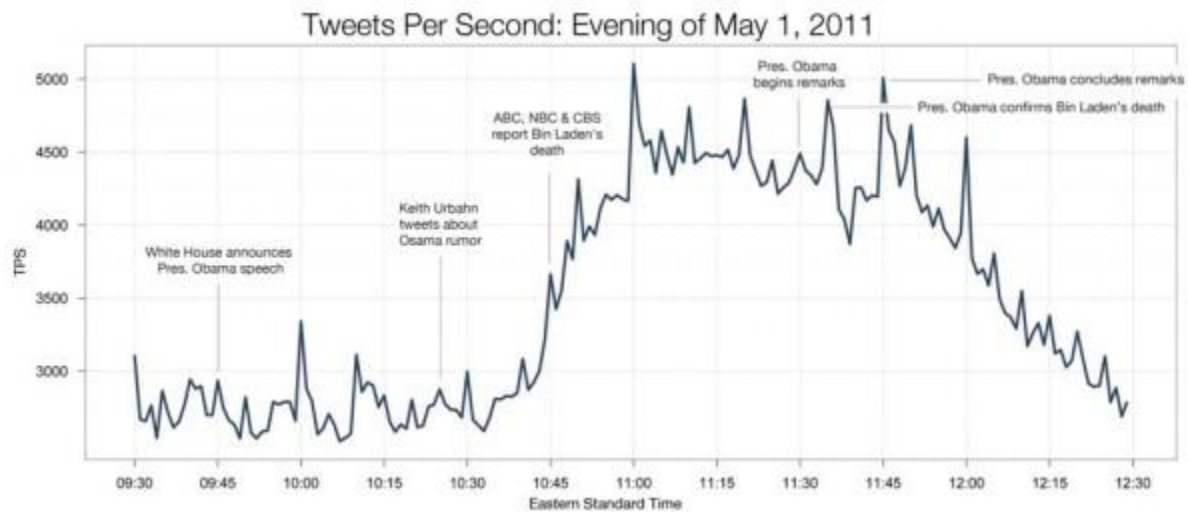
4,064 [FEB 6](#) Superbowl

5,530 [MAR 11](#) Japanese earthquake and tsunami

3,966 [APR 29](#) UK Royal Wedding

5,106 [MAY 2](#) Raid on Osama bin Laden

6,303 [MAY 28](#) UEFA Champions League





M

Euro 2012 Finals: 15,358 TPS

http://mashable.com/2012/02/06/tweets-per-second-records-twitter/#PW_OA_1vfGqo



“ ” @twittercomms
Twitter Comms

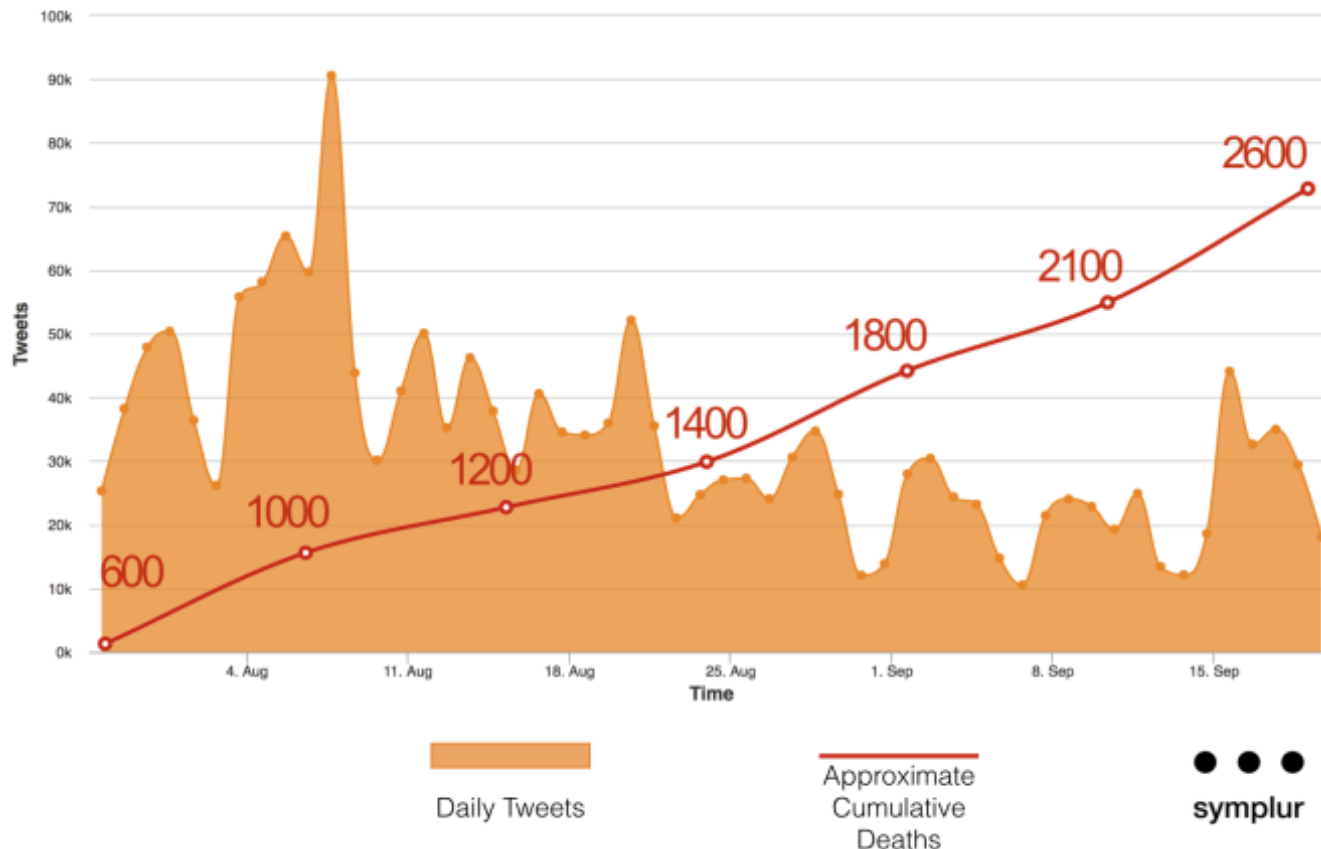
On Dec 9, the television screening in Japan of Hayao Miyazaki’s “Castle in the Sky” led to 25,088 Tweets per second - a new Twitter record.



The same peak detection analysis can be performed on individual topics


Ebola - Tweets decline while death toll rises

Twitter, Jul 28 through Sep 19, 2014: #Ebola, #EbolaOutbreak, #EbolaVirus, #EbolaWatch



Baseline

- Baseline in 2013: 6000 TPS



New Tweets per second (TPS) record: 143,199
TPS. Typical day: more than 500 million Tweets
sent; average 5,700 TPS.

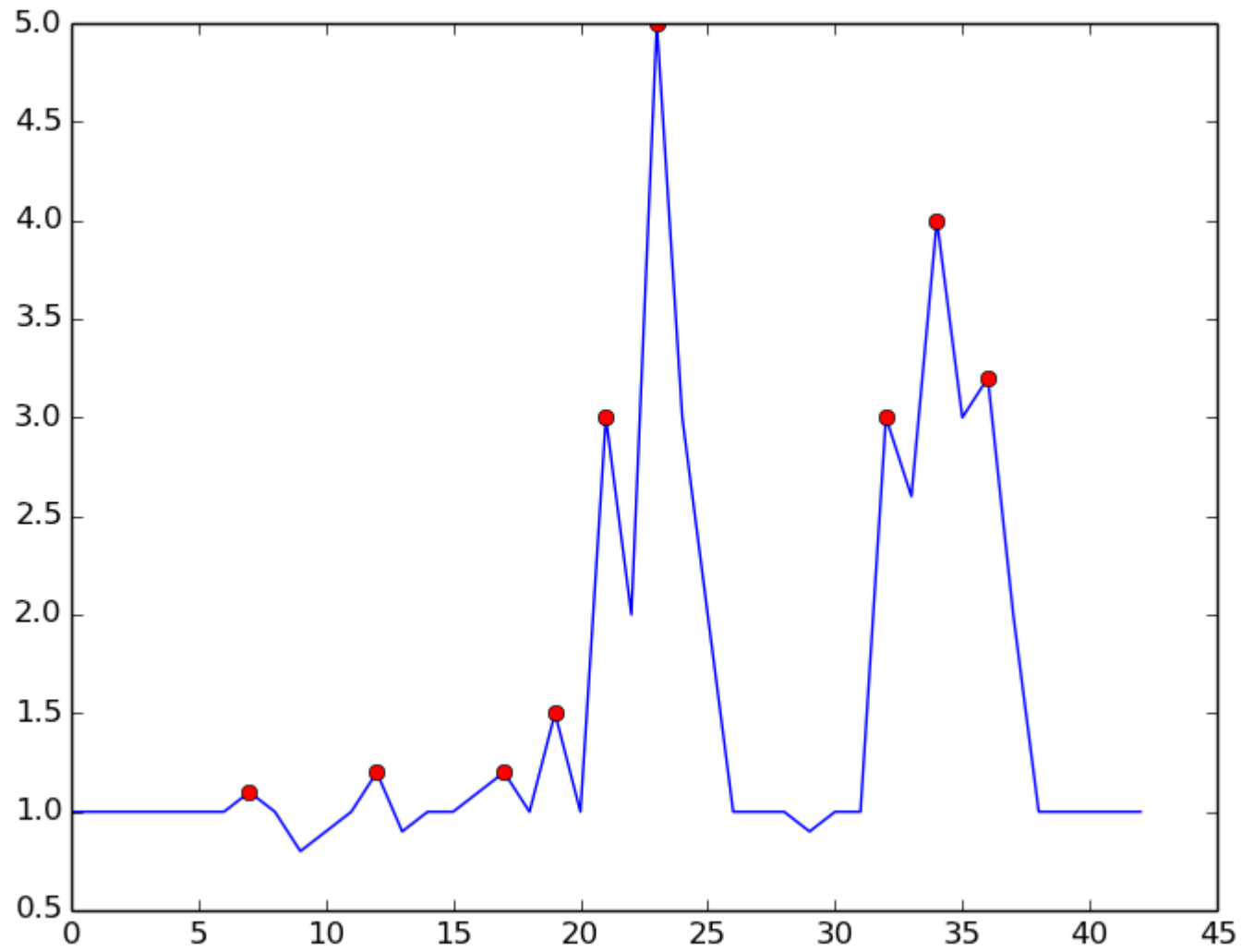
[Tweet](#)

What does the data look like?

Time Stamp	TPS
2 nd Feb 9:13:58	3123
2 nd Feb 9:13:59	2871
2 nd Feb 9:14:00	4509
2 nd Feb 9:14:01	8773
2 nd Feb 9:14:02	3470
2 nd Feb 9:14:03	1456
2 nd Feb 9:14:04	4873
2 nd Feb 9:14:05	3832
2 nd Feb 9:14:06	2412
2 nd Feb 9:14:07	4006

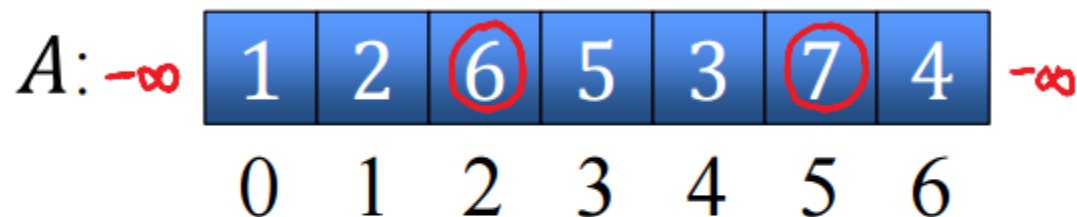
**What if we just took the top
100 TPS values?
Would that work?**

Time t is a peak if $(y(t) > y(t-1)) \ \&\& \ (y(t) > y(t+1))$



peak if $(y(t) - y(t-dt) > m) \ \&\& \ (y(t) - y(t+dt) > m)$

- Given an array $A[0..n-1]$:



- $A[i]$ is a **peak** if it is not smaller than its neighbor(s):

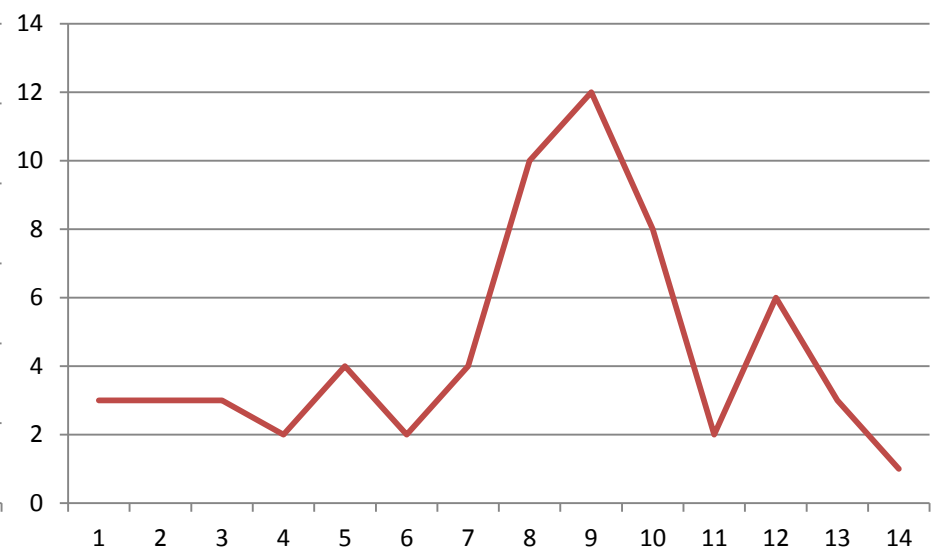
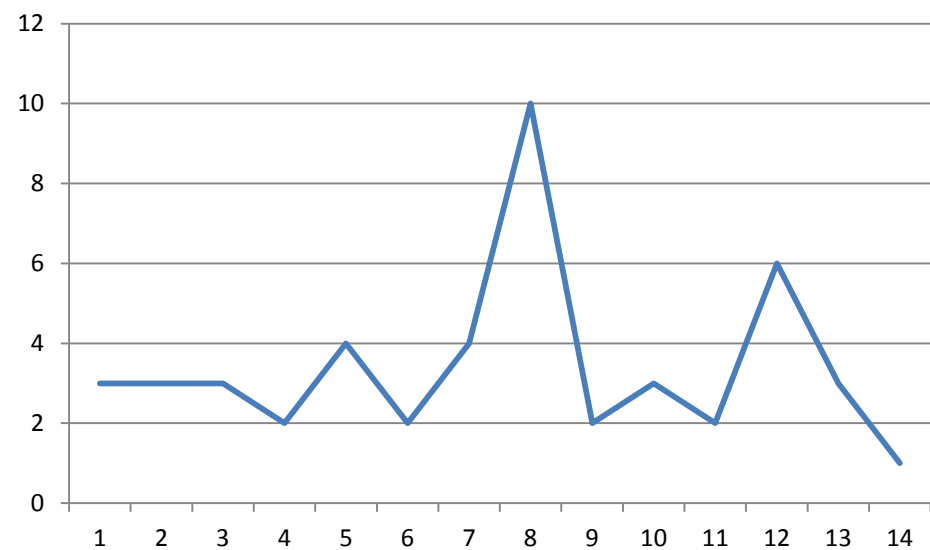
$$A[i-1] \leq A[i] \geq A[i+1]$$

But How do we define a Peak?

Is it different from sustained activity?

How to handle Prolonged Peaks

- A one-second peak, vs. (say) a 3-second peak



Cleverer Implementation

- Look at any element $A[i]$ and its neighbors $A[i - 1]$ & $A[i + 1]$
 - If peak: return i
 - Otherwise: locally rising on some side



We can start building “features” (new columns)

Time Stamp	TPS	Diff 1	Diff 2	Diff 3
2 nd Feb 9:13:58	3123	2871-3123	4509 - 3123	8773- 3123
2 nd Feb 9:13:59	2871			
2 nd Feb 9:14:00	4509			
2 nd Feb 9:14:01	8773			
2 nd Feb 9:14:02	3470			
2 nd Feb 9:14:03	1456			
2 nd Feb 9:14:04	4873			
2 nd Feb 9:14:05	3832			NA
2 nd Feb 9:14:06	2412	4006 - 2412	NA	NA
2 nd Feb 9:14:07	4006	NA	NA	NA

Now let's look at Time Series Data