"Eye of the Storm": Social Sensing of Extreme Weather Events Using Social Media

Michelle Spruce
PhD Computer Science

PI – Dr Hywel T. P. Williams

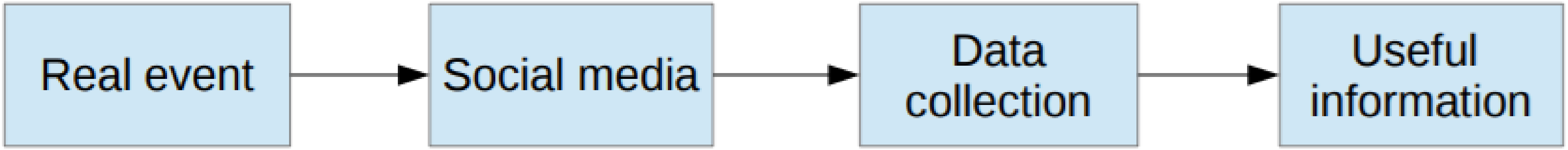
Research undertaken as part of University of Exeter SEDA Lab (Social & Environmental Data Analysis @ University of Exeter)

Project Context

- **Extreme weather events** cause disruption to communities and economies.
- However the **specific impacts** can be hard to forecast and observe.
- **‘Social Sensing’** provides an opportunity to improve understanding of impacts of extreme weather events.
- With **Social Media**, public can comment on and respond to their experiences of events such as extreme weather events.
- **Improved understanding** of impacts would allow better verification of meteorological forecast models and **aid impact based forecasting**

Social Sensing

- The systematic analysis of unsolicited social media data (*user generated content shared socially via the web*) to observe real-world events
- Involves event detection, location and characterisation



Aim:

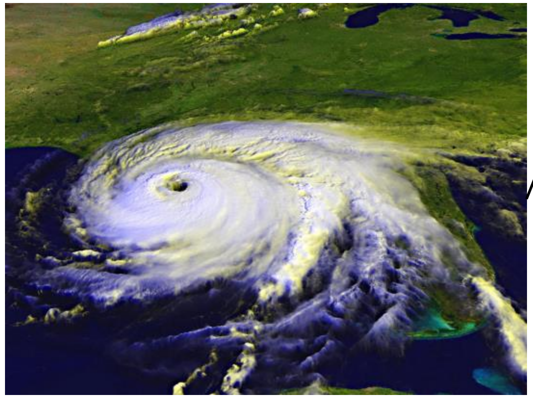
- To use ‘Social Sensing’ to **map** extreme weather events and **understand the social impacts**

Questions:

- What social impact information can we determine from social media?
- Can we assess the impact severity of an extreme weather event using Social Media?

Wildfires
Boulton et al, 2016

Earthquakes
Sakaki et al, 2010

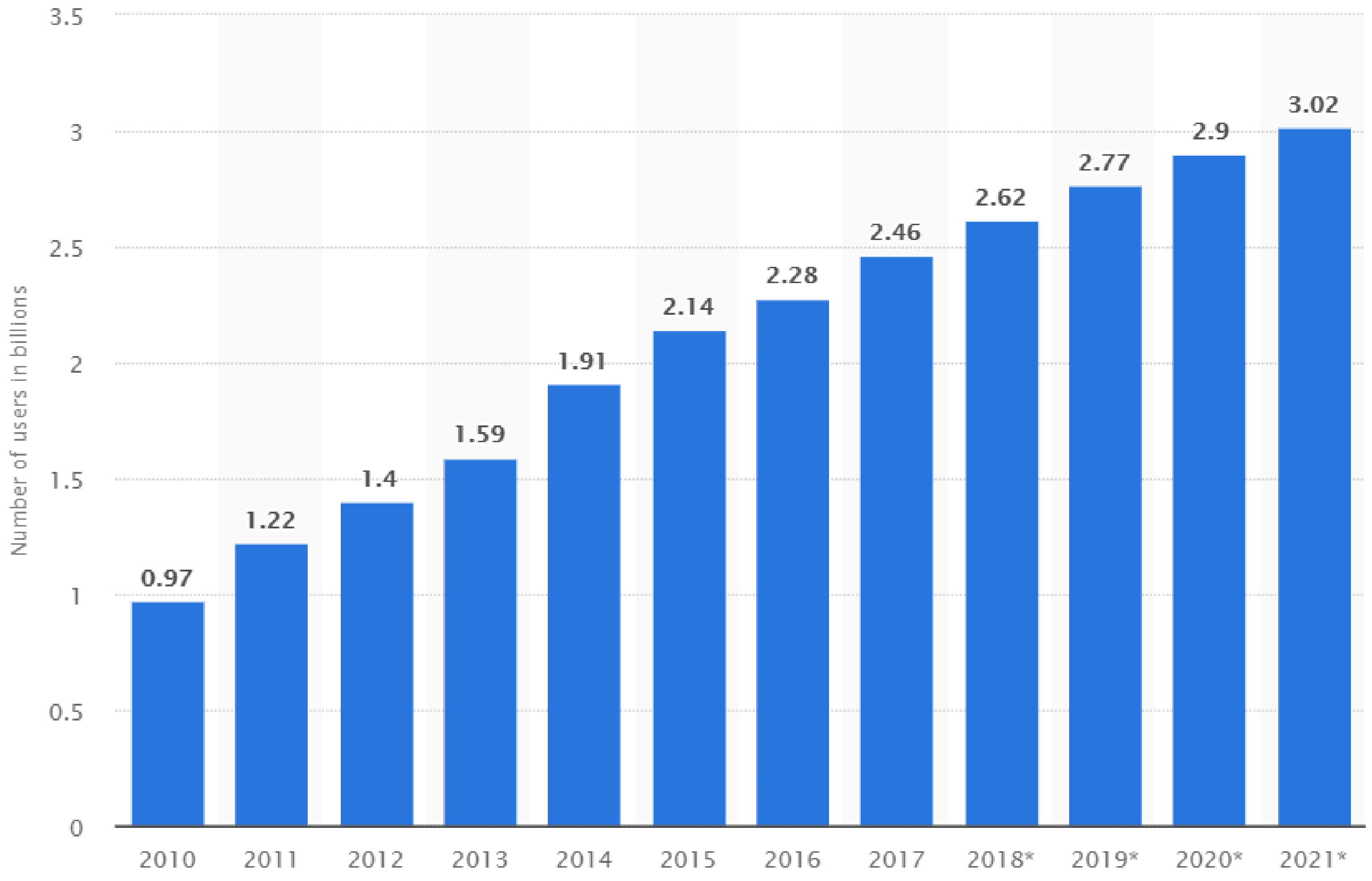


Hurricanes

Floods / Heavy Rain
Arthur et al, 2018

Heatwaves

Number of social media users worldwide from 2010 to 2021 (in billions)



https://www.statista.com

Tue 3rd July 11:30am
British Summer Time (BST)

Searching for floods on Twitter…

Connor @ConnorFlood30 · 5h
damn this weather is hella nice
3 4

Joyce Peterson @MemphoNewsLady · 6h
#memsunset pics 4 days apart at Greenbelt Park in Harbor Town.
#MississippiRiver at Memphis is about to hit 34ft **flood** stage. (33.6ft now.)

2 5 16

Fawks @OhFawker · May 7
This area is normally a field ... Water is almost up to the bench seats

#OttawaGatineauFlood... instagram.com/p/BTzHiEkhWzq/

Natchez Events @NatchezEvents · 8m
Officials keep eyes on Mississippi River as it rises above **flood** stage ift.tt/2ql2a2G

Ottawa Citizen @OttawaCitizen · 6h
Photos: The view from the sky of the massive Ottawa-Gatineau **flood** ow.ly/No4W30bybxV

24 17

...or for news about Storm Hector...

Dougal @DougalCMK Follow

Storm Hector has already delivered me a few presents. So far a plant pot, a garden glove, couple empty bags of compost and a cat bed. Have it on good authority that the trampoline may be on it's way 😵💨 #stormhector

6:39 AM - 14 Jun 2018

1 Like

1

Sharon Mohamed @SharonMohamed3 Follow

Thank you storm hector... it appears I have a heli pad now instead of a trampoline 😂😂

10:34 AM - 14 Jun 2018

Metro @MetroUK Follow

Trampolines go rogue as Storm Hector smashes into Britain with 100mph winds

Trampolines go rogue as Storm Hector smashes into Britain with 100mph win...

People are being advised to tie down their garden furniture and trampolines to avoid disruption during rush hour.

metro.co.uk

9:10 AM - 14 Jun 2018

3 Retweets 5 Likes

1 3 5

MayoRoadsafety @safecounty Follow

Good morning, #mayo It's a wild one out there today with some debris on the road. Storm Hector has arrived and has brought strong gale force winds, so be careful. Take care and remember to tie those trampolines down .@MayoDotIE .@MayoCoCo

9:27 AM - 14 Jun 2018

Tom @tomstyleo Follow

Give over... its not 'Storm Hector' its just a bit blowy

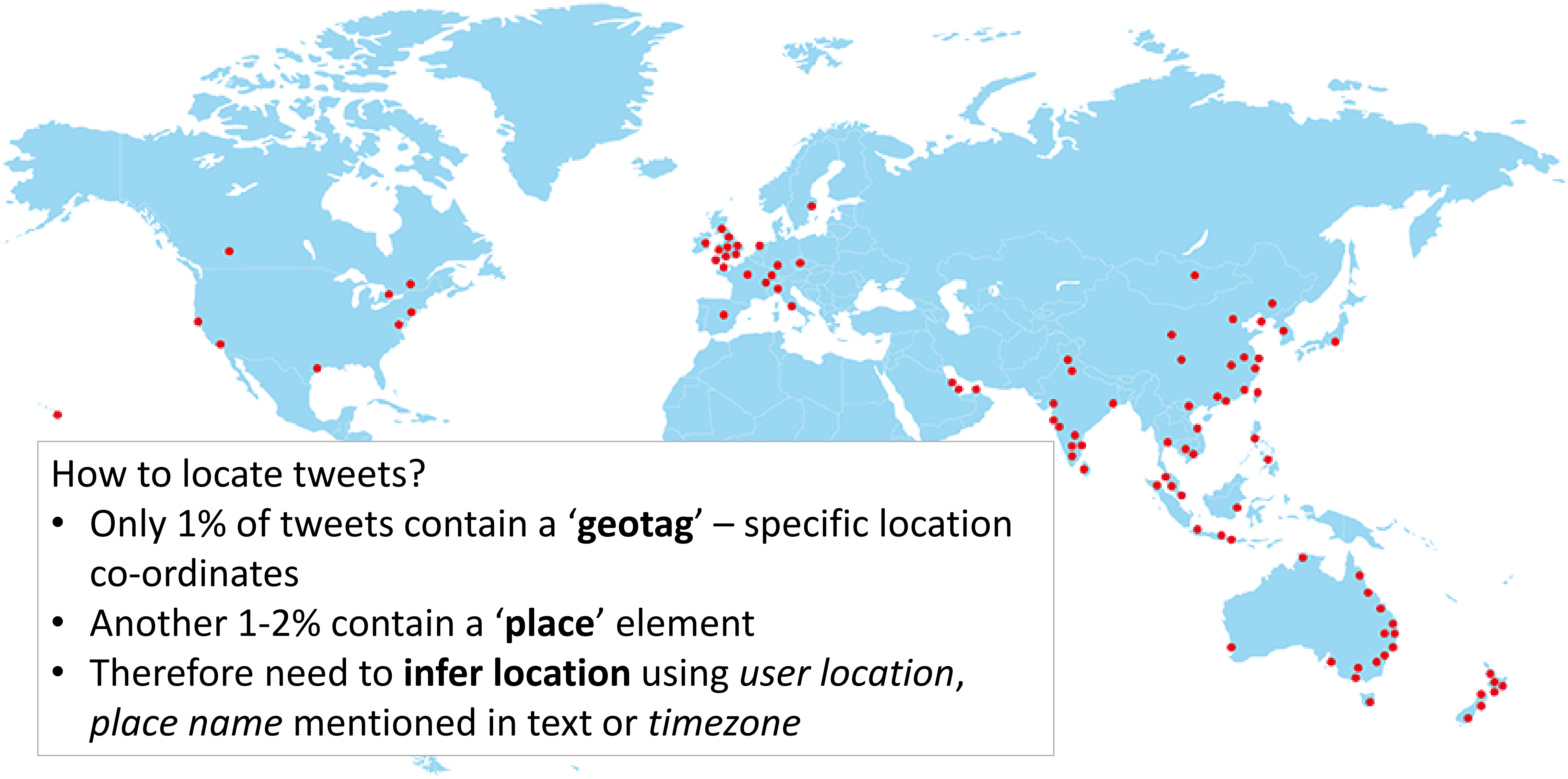
Weather is just weather.

10:30 AM - 14 Jun 2018 from North East, England

Aidan Comerford @MrAComerford · Jun 13

After **Storm** Doris, we had to apologise for our trampoline in our neighbour's garden. Their kid thought it was a present, cos it was her birthday. To make matters worse, we asked if the trampoline had hurt their dog only to be told that he died at Xmas.#badneighbours #StormHector

12 84 627



How to locate tweets?

- Only 1% of tweets contain a ‘**geotag**’ – specific location co-ordinates
- Another 1-2% contain a ‘**place**’ element
- Therefore need to **infer location** using *user location*, *place name* mentioned in text or *timezone*

PLOS | ONE

RESEARCH ARTICLE

Social sensing of floods in the UK

Rudy Arthur[1]*, **Chris A. Boulton**[2☯], **Humphrey Shotton**[1☯], **Hywel T. P. Williams**[1]

1 Computer Science, CEMPS, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, United Kingdom, **2** Earth System Science, CLES, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, United Kingdom

☯ These authors contributed equally to this work.
* R.Arthur@exeter.ac.uk

Check for updates

Citation: Arthur R, Boulton CA, Shotton H, Williams HTP (2018) Social sensing of floods in the UK. PLoS ONE 13(1): e0189327. https://doi.org/10.1371/journal.pone.0189327

Editor: Guy J-P. Schumann, Bristol University/ Remote Sensing Solutions Inc., UNITED STATES

Received: July 11, 2017

Accepted: November 22, 2017

Published: January 31, 2018

Abstract

"Social sensing" is a form of crowd-sourcing that involves systematic analysis of digital communications to detect real-world events. Here we consider the use of social sensing for observing natural hazards. In particular, we present a case study that uses data from a popular social media platform (Twitter) to detect and locate flood events in the UK. In order to improve data quality we apply a number of filters (timezone, simple text filters and a naive Bayes 'relevance' filter) to the data. We then use place names in the user profile and message text to infer the location of the tweets. These two steps remove most of the irrelevant tweets and yield orders of magnitude more located tweets than we have by relying on geo-tagged data. We demonstrate that high resolution social sensing of floods is feasible and we can produce high-quality historical and real-time maps of floods using Twitter.

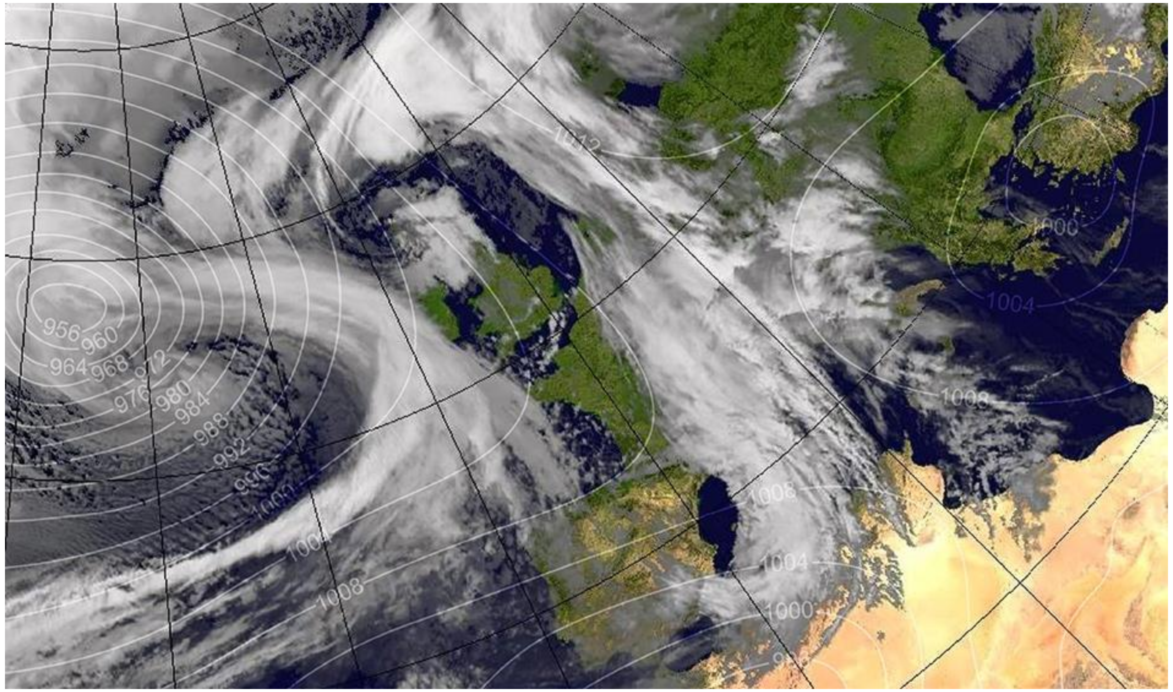
Introduction

Natural hazards such as floods, wildfires, storms and other extreme weather events cause substantial disruption to human activity and are predicted to increase in frequency and severity as

UK Storm Names 2017/18

Met Office

Name	Date named	Date of impact on UK and/or Ireland
Aileen	12 September 2017	12 - 13 September 2017
Ex-Hurricane Ophelia	11 October 2017 (Named by NHC)	16 - 17 October 2017
Brian	19 October 2017	21 October 2017
Caroline	5 December 2017	7 December 2017
Dylan	29 December 2017	30 - 31 December 2017
Eleanor	1 January 2018	2 - 3 January 2018
Fionn (F-yunn)	16 January 2018	16 January 2018
David	17 January 2018 (Named by Meteo France)	18 January 2018
Georgina	23 January 2018	24 January 2018
Hector	13 June 2018	13 - 14 June 2018



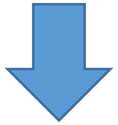
https://www.metoffice.gov.uk/barometer/uk-storm-centre

Collecting Twitter Data (Tweets)

Twitter collections – Using Twitter API[1] collect tweets containing natural hazard/weather related key words – tweets received in JSON format

Storm Name Collection: *storm, Ophelia, Ofelia, Opelia, Ohpelia, ophelia, Opehlia, opheliaireland, Brian, brian, Caroline, caroline, Dylan, dylan, Eleanor, eleanor, Fionn, fionn, Fion, fion, Georgina, georgina, Emma, emma*

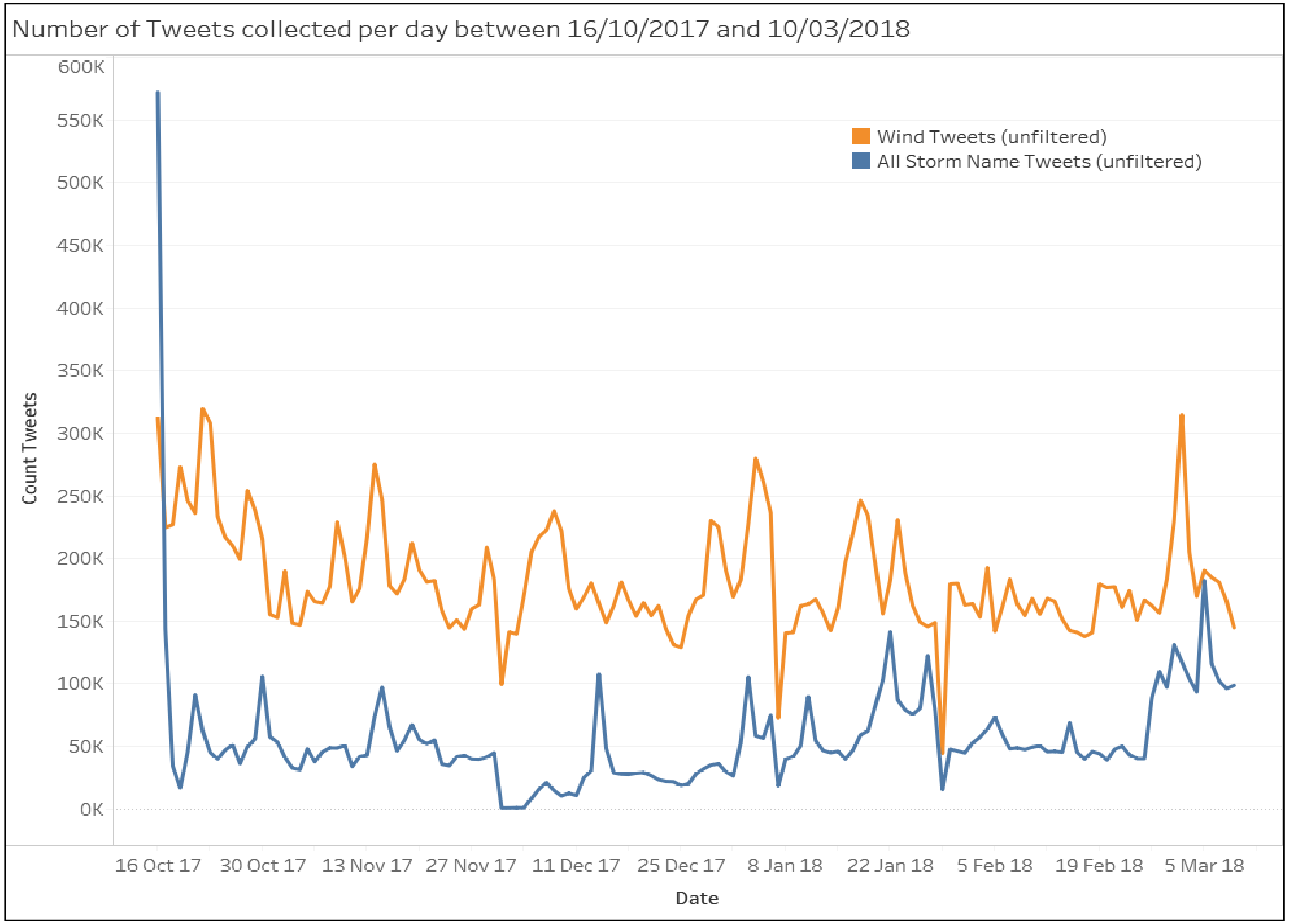
Wind Collection: *wind, gale, windstorm, hurricane*



Extract tweets from collections for a specific time period/keywords

https://developer.twitter.com/content/developer-twitter/en.html

- Tweets collected 16th October 2017 – 10th March 2018
- Over 35 million individual Tweets collected



Method – relevance filter

Bot filter

- Remove usernames known to be automated 'Twitter bots'

Weather station filter

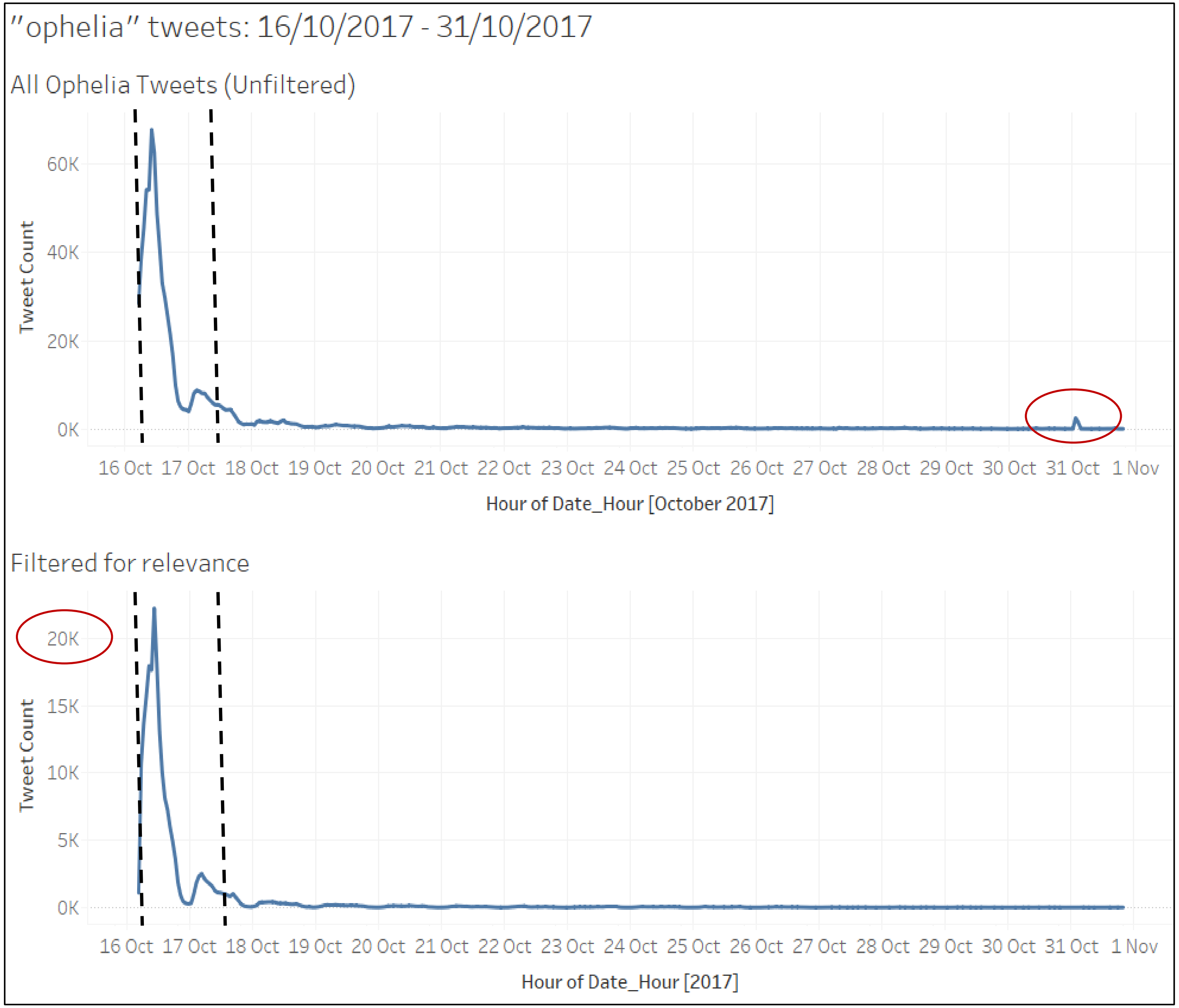
- *Remove tweets with known automated weather station structure*

Irrelevant Term filter

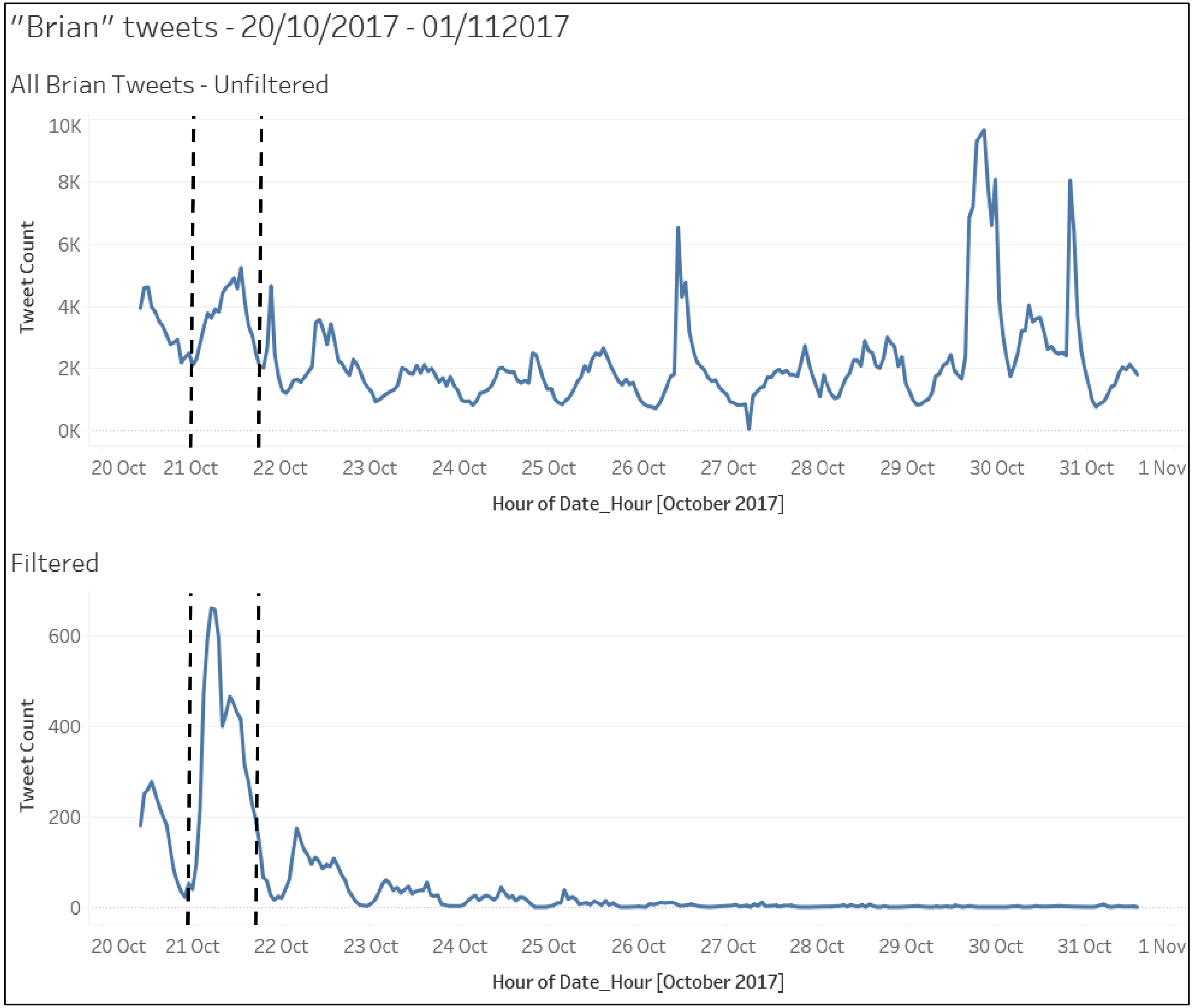
- *Exclude tweets containing terms such as: 'cook up a storm', 'wind up', 'throw caution to the wind'*, etc)

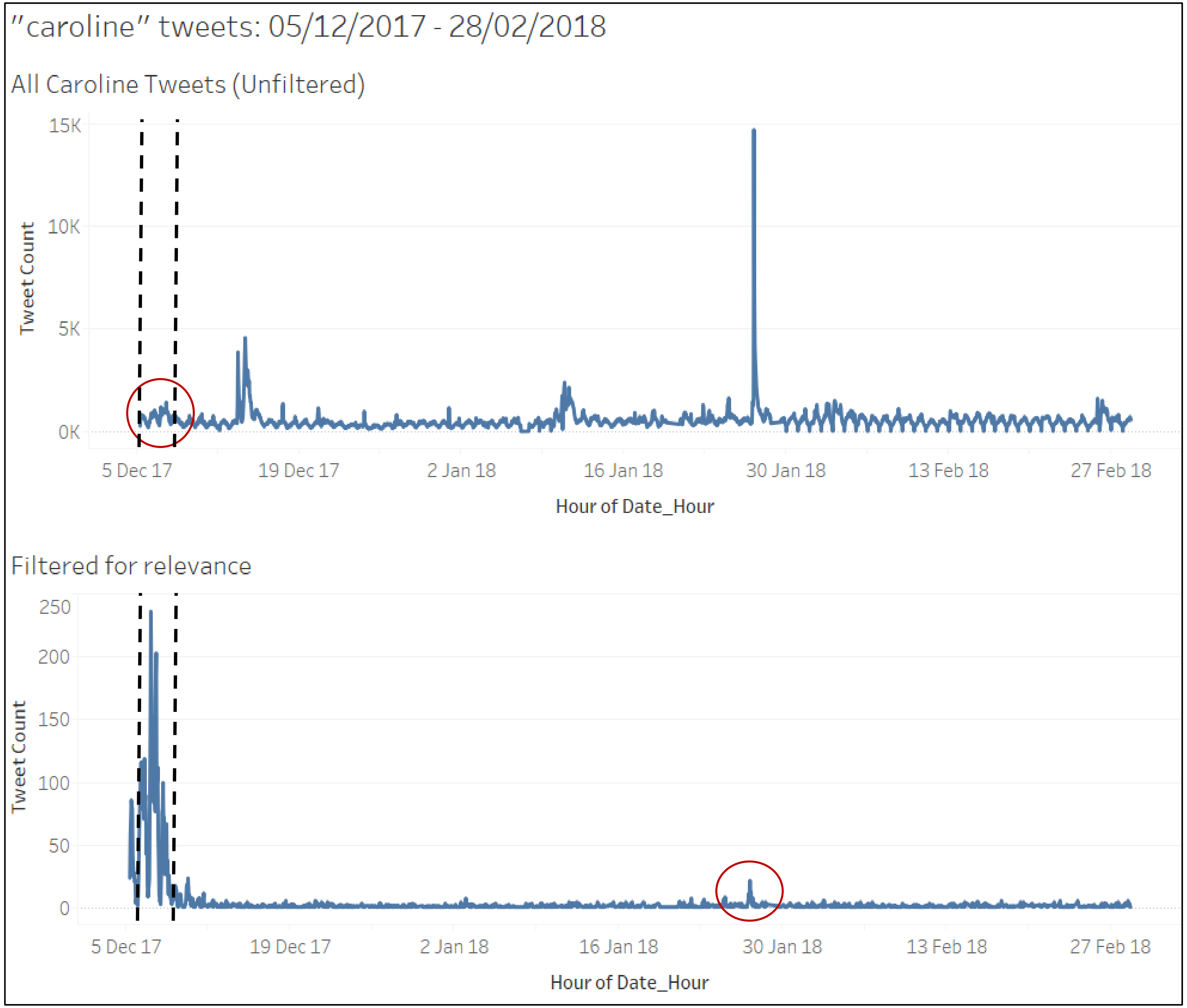
Machine learning

- *Created training corpus using 5000 labelled tweet examples*
- *Apply **Naïve Bayes** algorithm to remove tweets based on training corpus*



"ophelia" tweets: 16/10/2017 - 31/10/2017
All Ophelia Tweets (Unfiltered)
Tweet Count
60K
40K
20K
0K
16 Oct 17 Oct 18 Oct 19 Oct 20 Oct 21 Oct 22 Oct 23 Oct 24 Oct 25 Oct 26 Oct 27 Oct 28 Oct 29 Oct 30 Oct 31 Oct 1 Nov
Hour of Date_Hour [October 2017]
Filtered for relevance
Tweet Count
20K
15K
10K
5K
0K
16 Oct 17 Oct 18 Oct 19 Oct 20 Oct 21 Oct 22 Oct 23 Oct 24 Oct 25 Oct 26 Oct 27 Oct 28 Oct 29 Oct 30 Oct 31 Oct 1 Nov
Hour of Date_Hour [2017]

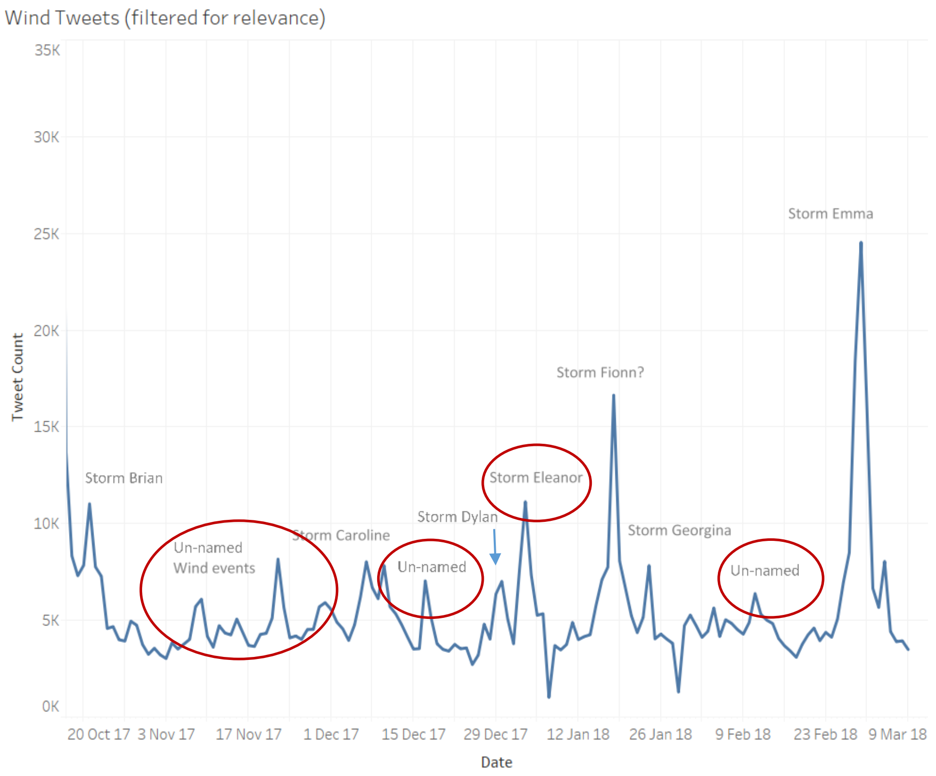

"Brian" tweets - 20/10/2017 - 01/112017
All Brian Tweets - Unfiltered
Tweet Count
10K
8K
6K
4K
2K
0K
20 Oct
21 Oct
22 Oct
23 Oct
24 Oct
25 Oct
26 Oct
27 Oct
28 Oct
29 Oct
30 Oct
31 Oct
1 Nov
Hour of Date_Hour [October 2017]
Filtered
Tweet Count
600
400
200
0
20 Oct
21 Oct
22 Oct
23 Oct
24 Oct
25 Oct
26 Oct
27 Oct
28 Oct
29 Oct
30 Oct
31 Oct
1 Nov
Hour of Date_Hour [October 2017]

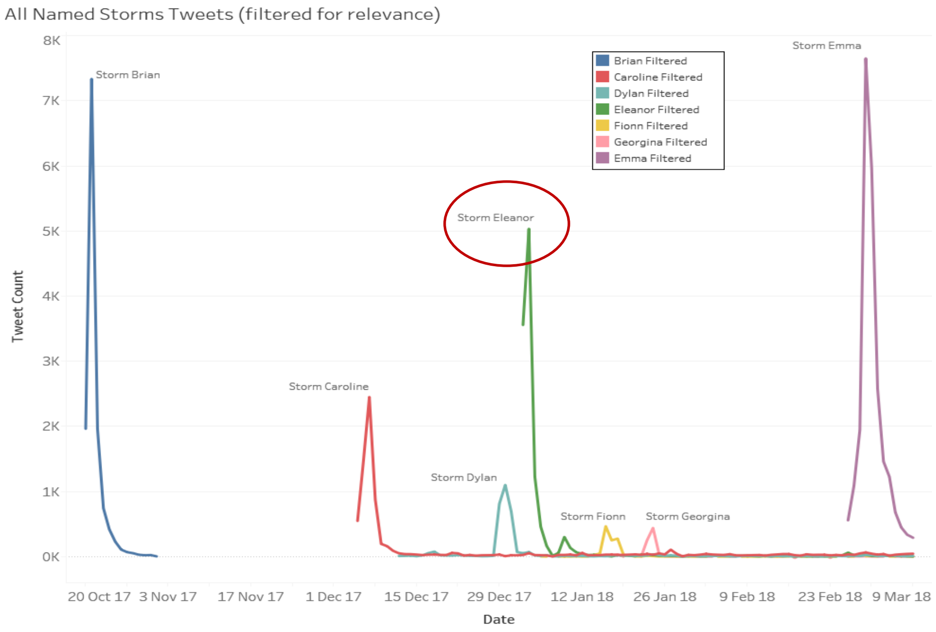

"caroline" tweets: 05/12/2017 - 28/02/2018
All Caroline Tweets (Unfiltered)
Tweet Count
15K
10K
5K
0K
5 Dec 17
19 Dec 17
2 Jan 18
16 Jan 18
30 Jan 18
13 Feb 18
27 Feb 18
Hour of Date_Hour
Filtered for relevance
Tweet Count
250
200
150
100
50
0
5 Dec 17
19 Dec 17
2 Jan 18
16 Jan 18
30 Jan 18
13 Feb 18
27 Feb 18
Hour of Date_Hour

Wind tweets (filtered for relevance)
versus
Named storm tweets (filtered for relevance)

- Peaks in wind tweets coincide with peaks in named storm tweets

- Also peaks in wind tweet activity at other times to named storm events – are these ‘extreme weather events’? Or just windy days?





Method – Location Inference

- Filter for **timezone** (UK/Ireland only)

- Locate tweets using **Geotag** (GPS cords)
  - **Place** (polygon cords)

- **User location** (*GPS cords, if not lookup text with Geonames*[1] *db*)

- **Place names** mentioned in tweet text (*dbpedia*[2] *spotlight lookup identifies placenames – then Geonames lookup to get coords*)

1 http://www.geonames.org/
2 https://wiki.dbpedia.org/

Tweet Collection	All Tweets (unfiltered)	Filtered for relevance	% of All Tweets	Filtered for relevance AND located	% of All Tweets	% of Filtered for relevance Tweets
1. Wind	26,298,449	831,076	3.2%	472,586	1.8%	56.9%
2. All Storm names	8,101901	278412	3.4%	214,220	2.6%	76.9%
ophelia	897,054	214,730	23.9%	167,369	18.7%	77.9%
brian	2,037,045	12,970	0.6%	9,439	0.5%	72.8%
caroline	1,199,149	8,552	0.7%	4,993	0.4%	58.4%
dylan	2,504,264	3,907	0.2%	2,410	0.1%	61.7%
eleanor	555,433	11,872	2.1%	9,761	1.8%	82.2%
fionn	43,936	1,260	2.9%	878	2.0%	69.7%
georgina	104,327	894	0.9%	650	0.6%	72.7%
emma	760,693	24,227	3.2%	18,720	2.5%	77.3%

Locating Tweets

Tweet Collection	All Tweets (unfiltered)	Filtered for relevance AND located	Geo co-ords		Place co-ords		User location (co-ords)		User location (resolvable place name)		Place name mentioned in text	
1. Wind	26,298,449	473,740	7,351	1.6%	18,539	3.9%	21,871	4.6%	361,156	76.2%	64,823	13.7%
2. All Storm names	8,101901	214,220	1349	0.6%	7169	3.3%	933	0.4%	159207	74.3%	45562	21.3%

Social Impact

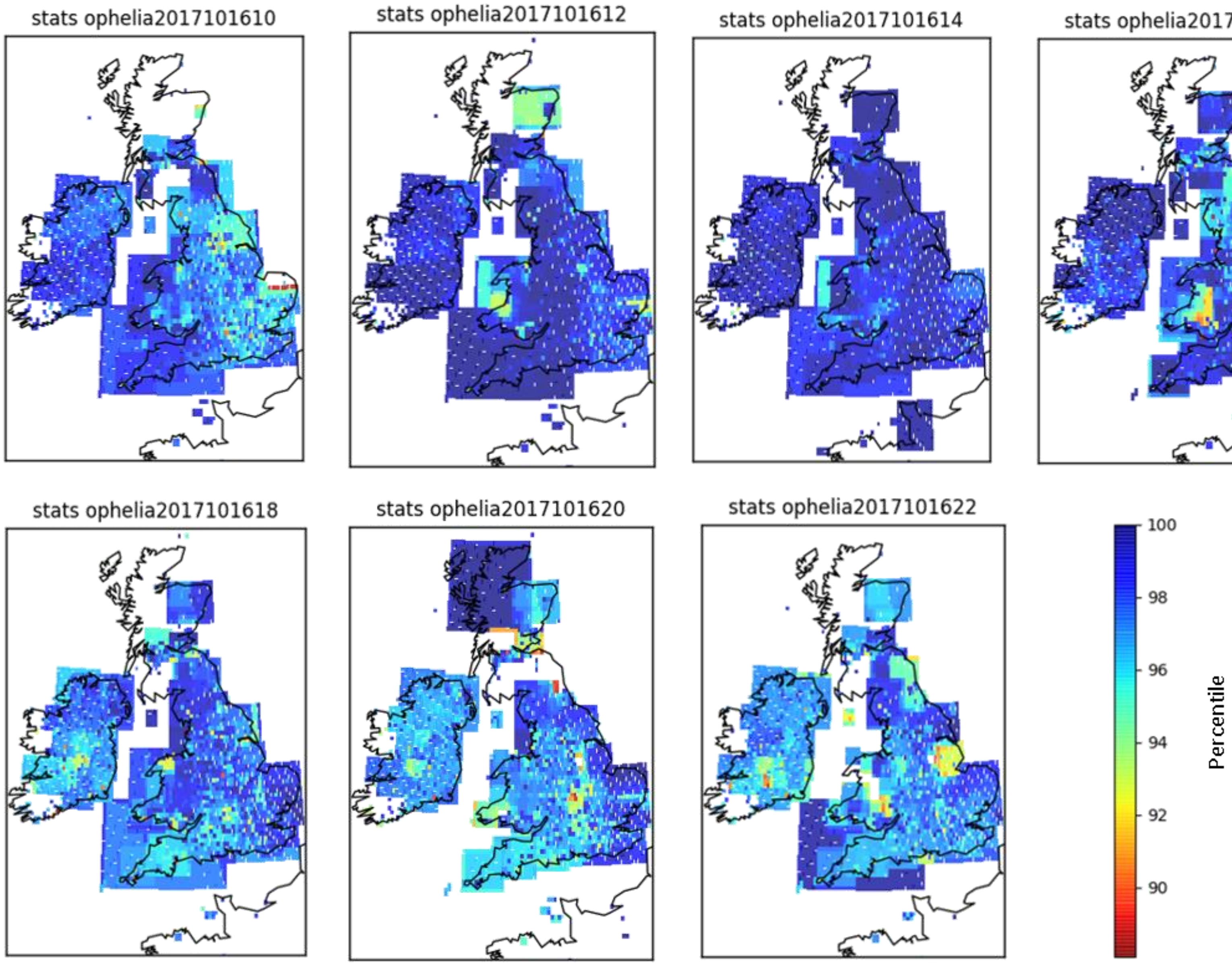
Using all filtered tweets:

- Count tweets by date/hour
- Sentiment analysis
- Categorisation (disruption, damage, warnings, news, humour, other)

Using located tweets:

- Map tweet activity (absolute, percentile, sentiment)
- Percentile normalises for variations in population, tweet activity, etc

Locating tweets during Ex-hurricane Ophelia – 16/10/2017



Sentiment Analysis

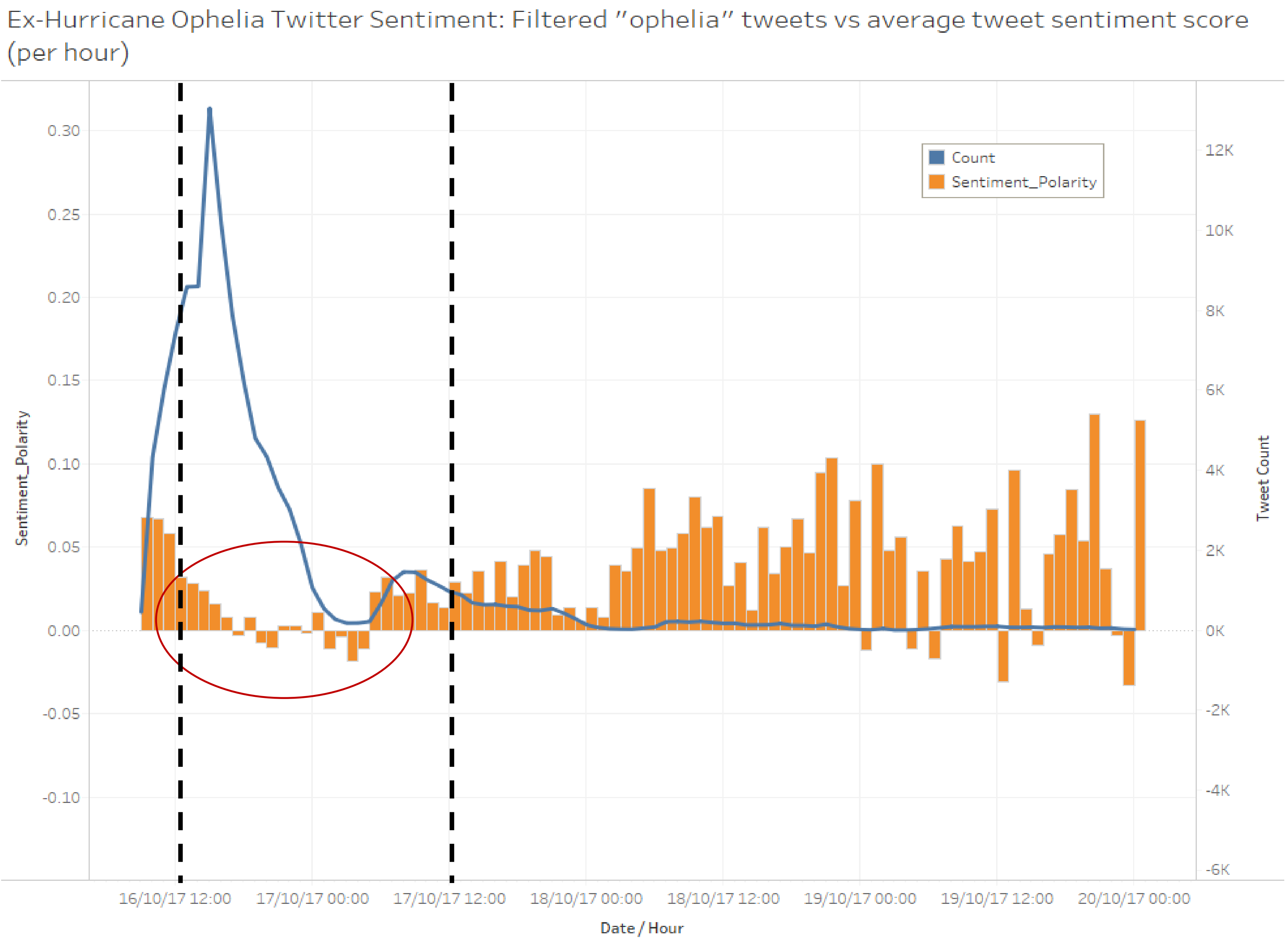
- Can we use social media posts to infer how a person feels about an event?

Machine Learning / Natural Language processing – use tweets as training data, assign tweet to sentiment

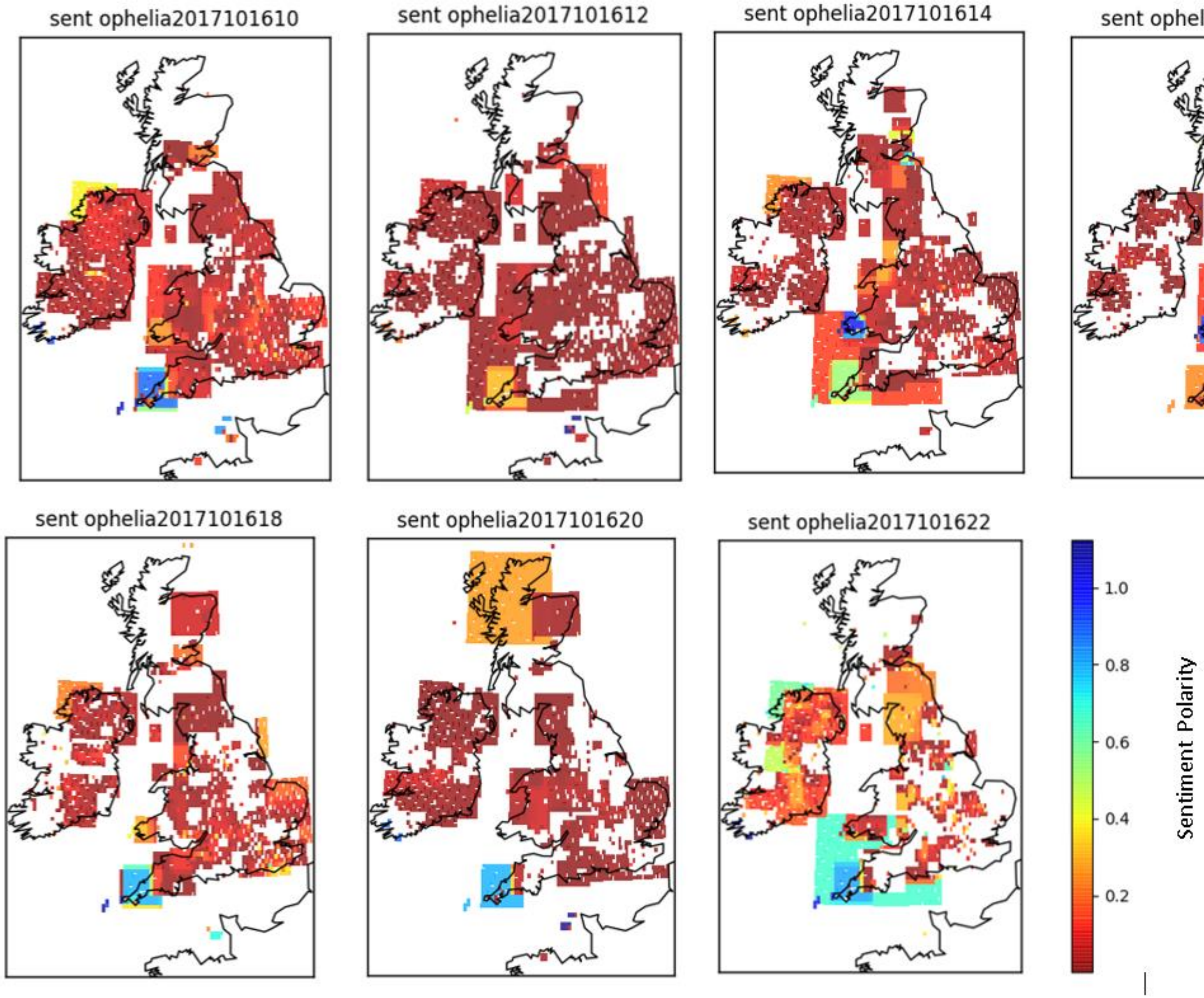
Lexicon based – look up to dictionary of sentiment words

Paltoglou G., Thelwall M. (2017) Sensing Social Media: A Range of Approaches for Sentiment Analysis. In: Holyst J. (eds) Cyberemotions. Understanding Complex Systems. Springer, Cham

Measuring the Sentiment of 'Ophelia'

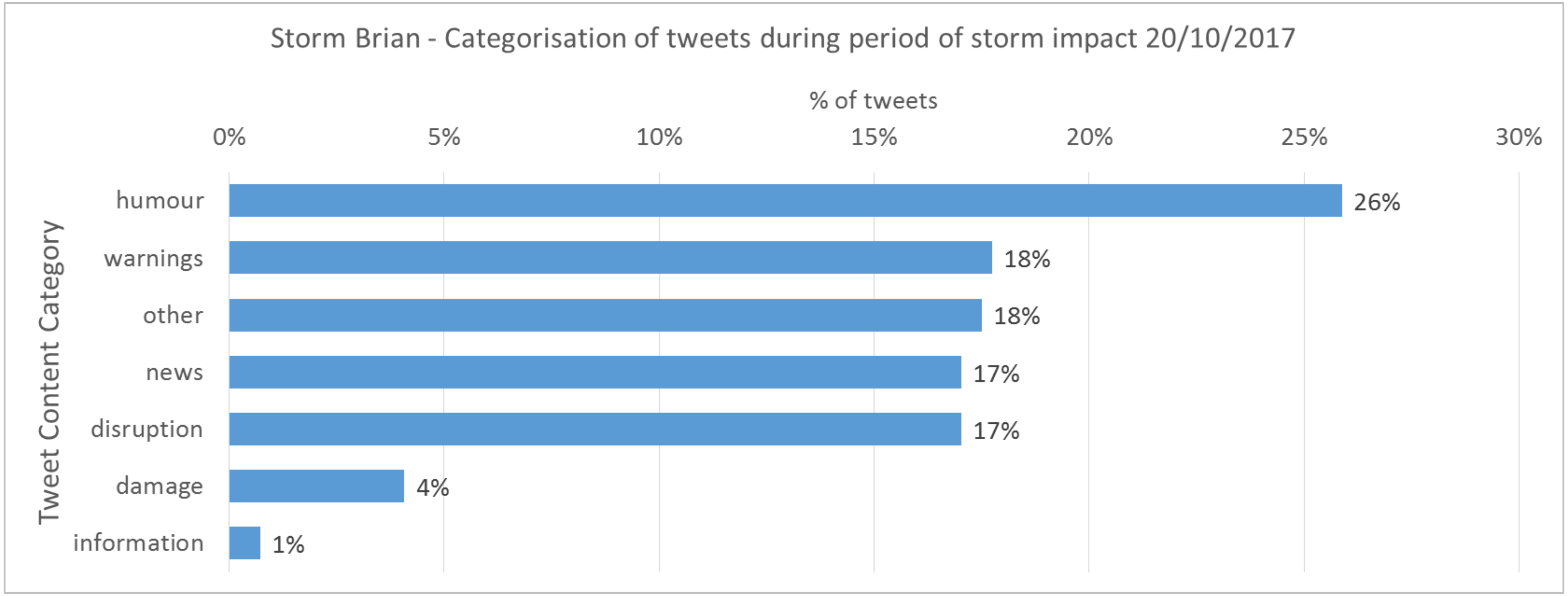


Tracking the Sentiment of 'Ophelia'



Categorisation of Tweets

Filtered tweets are manually labelled into categories:



Tweet Content – Storm Brian

(Example tweet text from each category, not actual Twitter posts)

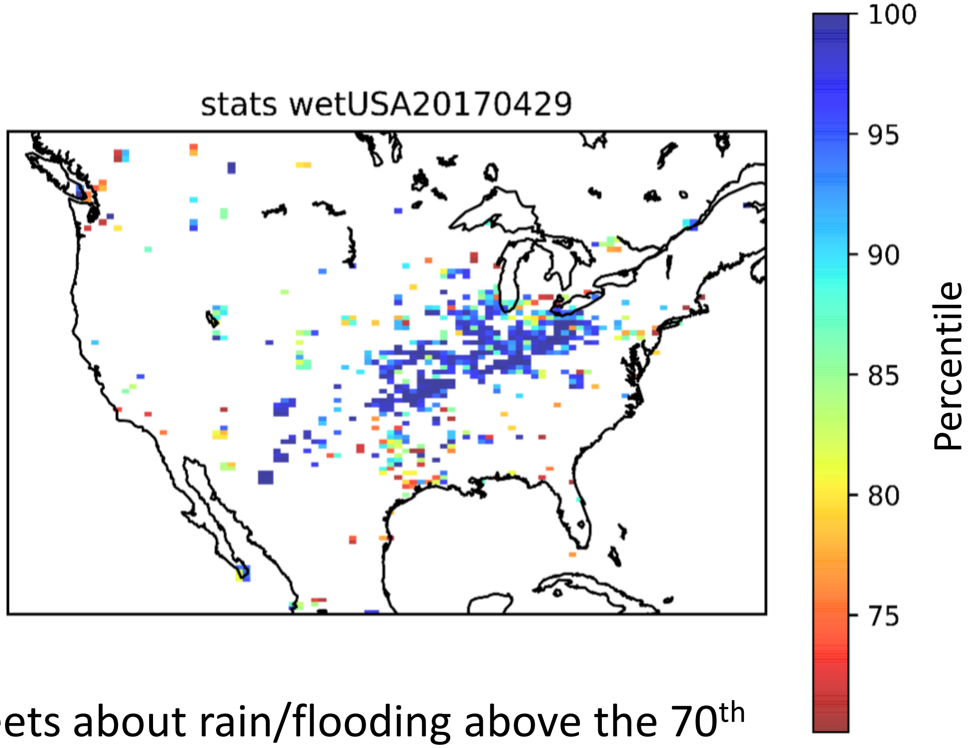
Humour	Disruption
'Brian? What kind of name is that for a storm? Everyone knows Brian is a snail.' *'Am I the only one to find it really hard to take a storm called #Brian seriously?'* *'And Brian? Really? Storm Rambo or Terminator would be far better than #StormBrian'*	*"Train delay: National Rail have warned of delays due to high winds from Storm Brian"* *"Storm Brian latest - tree blocks railway lines and hovercraft suspended"* *"Major motorway was CLOSED after Storm Brian floods carriageway"*
Warnings	**Damage**
'#StormBrian could lead to travel disruption this weekend.' *'Storm Brian set to batter UK with heavy rain and 70mph winds.'* *'Take care on the coast folks. Waves are quite high with #StormBrian'*	"This is the scene this morning as the waves have damaged the Harbour Office during Storm Brian." "Storm Brian damage causes floodlight damage. Revised home game vs @ChesterCityFC" "Scaffolding in Helsby High Street BLOWN OVER by #StormBrian high winds"

Next Steps…

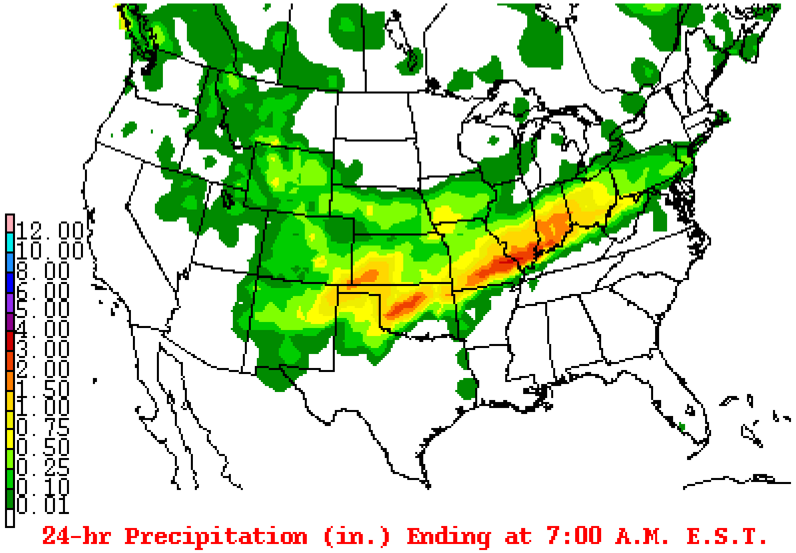
- (*in progress*) Use of Met Office observation / WOW data to compare 'Twitter talk' against observed weather conditions
- (*in progress*) Apply UK Storm model to extreme weather events outside of the UK
- Categorisation of tweets over time – how does this change over the period of the storm?
- Impact terminology – can we categorise the impact of an event using tweet text?

Applying model outside the UK

Searching for heavy rainfall and flooding…



Tweets about rain/flooding above the 70th percentile of these tweets for each grid square on 29th April 2017
Blue indicates unusually high number of tweets about rainfall/flooding



http://www.wpc.ncep.noaa.gov

Rainfall radar for the USA at 7am on 29th April 2017
Areas shaded red were most affected by heavy rainfall

Summary

- Social Sensing allows us to filter the ‘noise’ from Twitter and create a dataset of social media posts relating to extreme weather events
- 60-80% of tweets can be located using location inference method (as opposed to only 1% if just use geotag)
- Using percentile of tweets at a given time and place allows us to account for population size, prevalence of Twitter use, etc
- Sentiment analysis shows us that emotion becomes less positive during the period of a named storm
- Categorisation is a work in progress, however we find that about a quarter of tweets fall into the humour category
- **Overall social sensing provides a data source which can be used to help in understanding of the impact of weather events in addition to traditional methods**

Thank you

Michelle Spruce: m.spruce@exeter.ac.uk

Hywel Williams: h.t.p.williams@exeter.ac.uk

SEDA Lab: http://sites.exeter.ac.uk/seda-lab/