BD004 – Practical Text Analysis and Sentiment Analysis for Social Media

Dr Diana Maynard
University of Sheffield, UK

IADS Big Data and Analytics Summer School , 1 August 2017, University of Essex

Course Outline

- Introduction to social media analysis
- Adapting standard NLP tools for social media
- Introduction to sentiment analysis for social media
- Practical hands-on with sentiment analysis in GATE
  - Rule-based approaches
  - Machine-learning approaches
- Applications of sentiment and social media analysis
  - Political Futures Tracker
  - Analysing climate change tweets

1. Introduction to Social Media Analysis
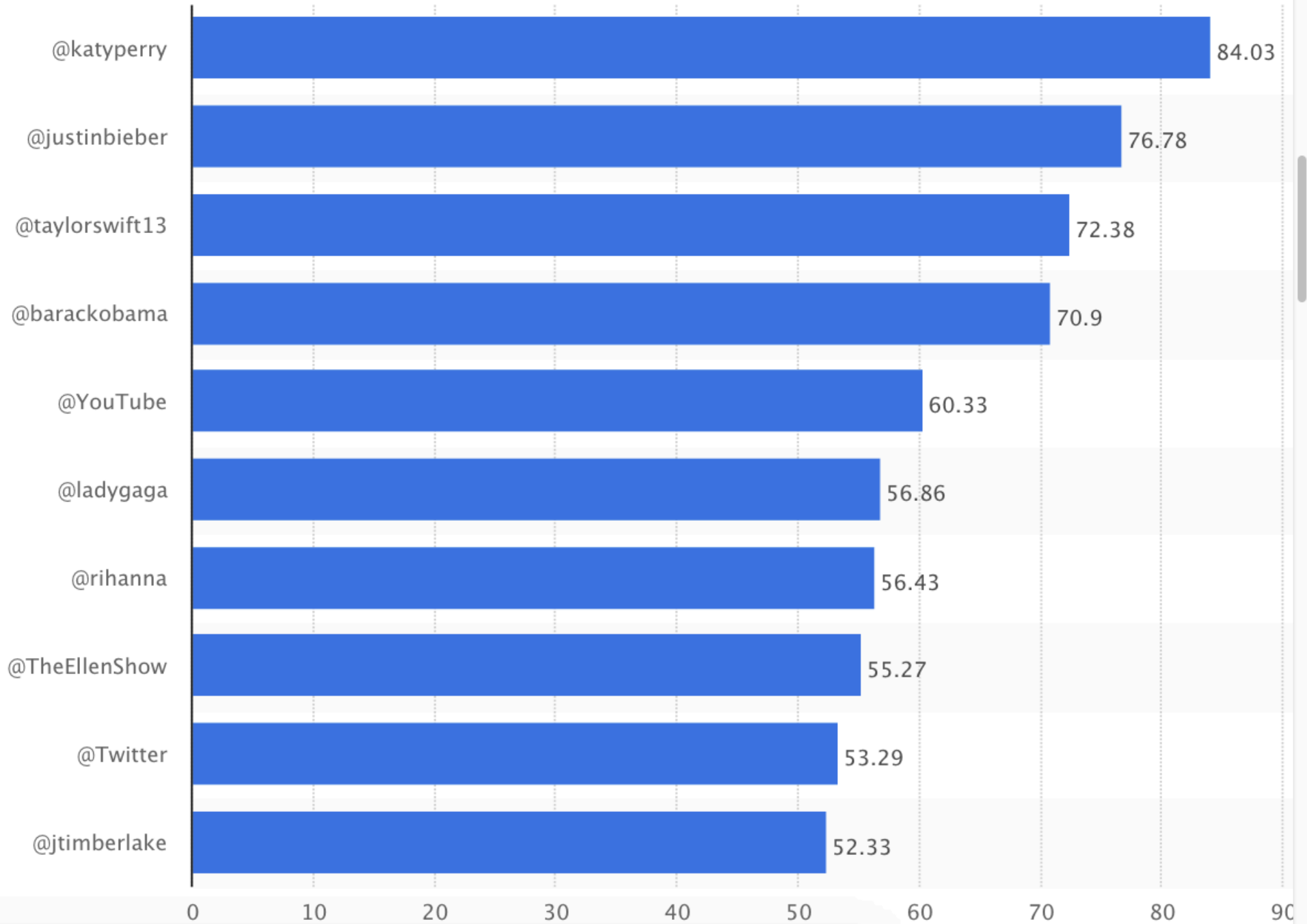
We are all connected to each other...
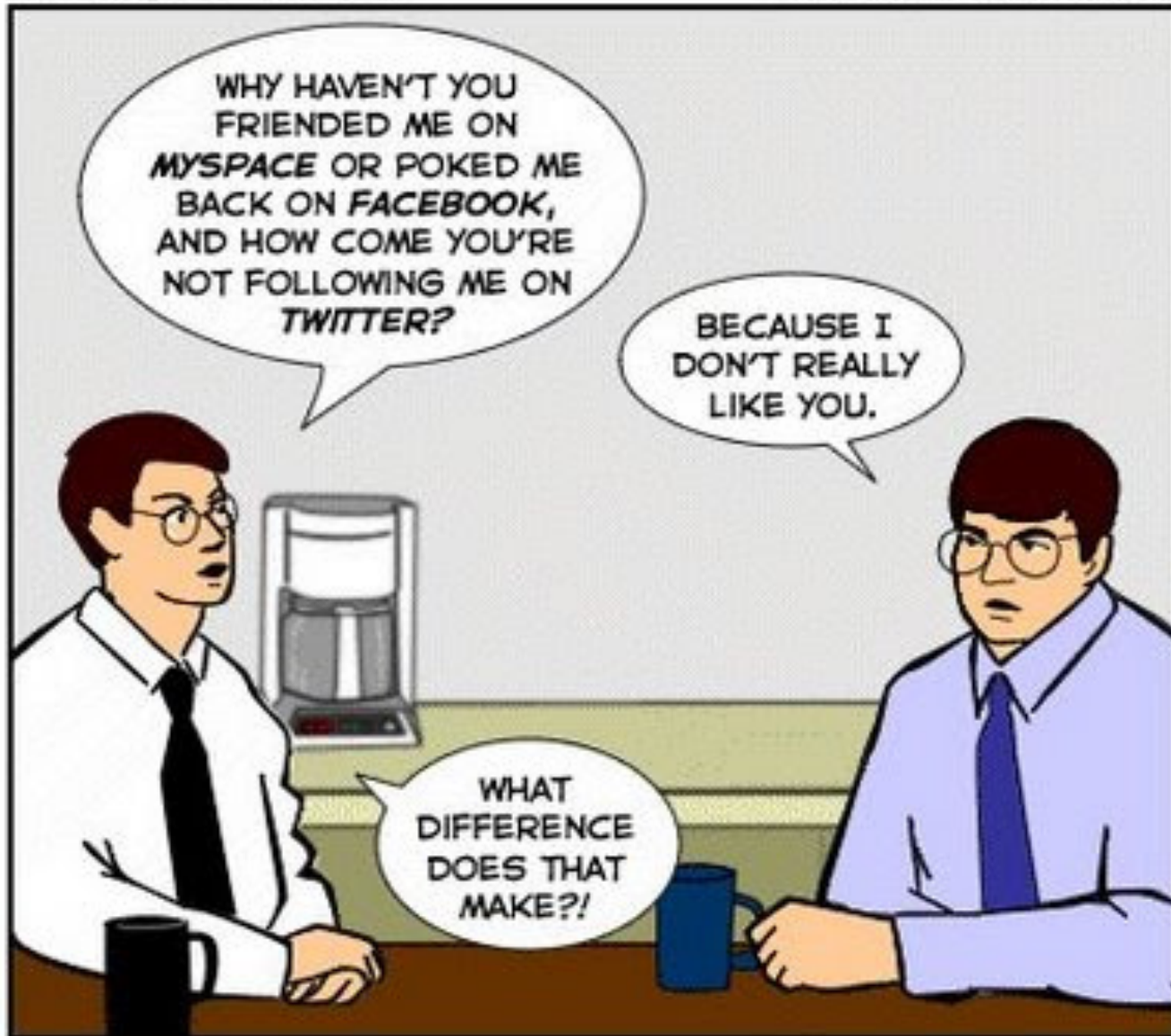
- Information, thoughts and opinions are shared prolifically on the social web these days
- 72% of online adults use social networking sites

Time Spend by Average Social networking user per month

facebook
405 minutes

Pinterest
405 mins

twitter
89 mins

Linked in
21 mins

Google+
3 mins

- In Britain and the US, we spend more than 1 hour a day on social media

- 90% of businesses use social media channels

Top 10 worldwide Twitter users

| User	Value
@katyperry	84.03
@justinbieber	76.78
@taylorswift13	72.38
@barackobama	70.9
@YouTube	60.33
@ladygaga	56.86
@rihanna	56.43
@TheEllenShow	55.27
@Twitter	53.29
@jtimberlake	52.33

Social Media and Big Data

- Big data is one of the biggest buzzwords of this decade
- If only we can harness all the information contained in social media, we have the **Holy Grail to business intelligence**
- There are already lots of examples of big data and social media analysis being used in real life…

Purchasing Recommendations

- Amazon recommends books (and other products) based on your previous purchases
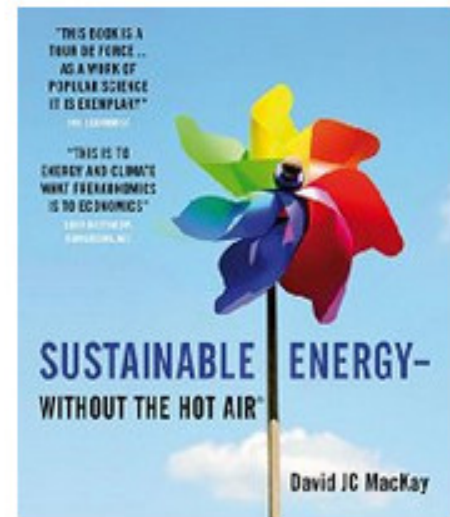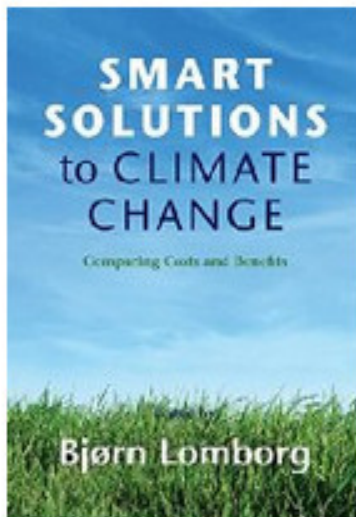
Sensor-driven big data

- Yarra trams in Melbourne, Australia, is using big data, sensor-driven data and the Internet of Things to deliver reports to central operations, and real-time alerts to maintenance techs and passengers.

Mood of the Nation

Mood Changes in UK Twitter Content 2009-2012

2010/10/30:
anger:0.48
fear:1.02
joy:0.21
sadness :0.5
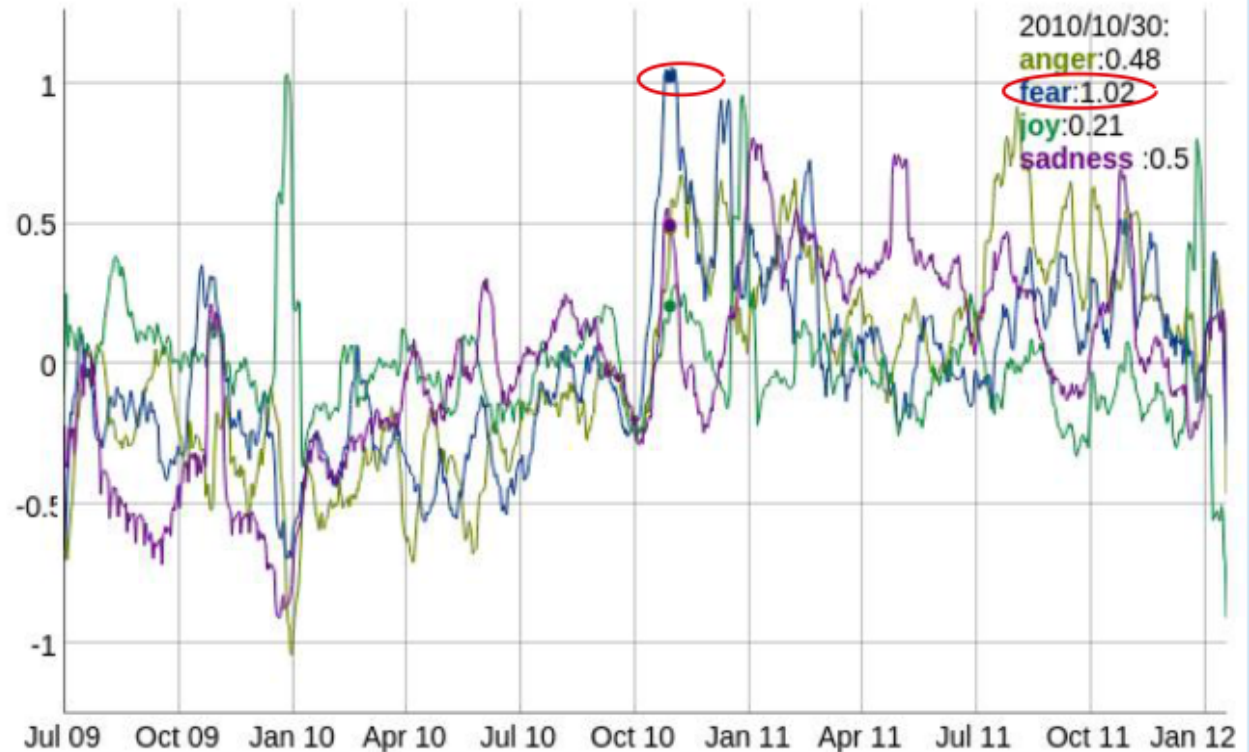
mediapatterns.enm.bris.ac.uk
Intelligent Systems Laboratory - University of Bristol

Social Media and the Stock Market

- Analysing public mood has proved useful for things like stock market prediction
- Fluctuations are driven mainly by fear rather than by things like happiness or sadness
- Although this could go horribly wrong...

IT'S THE PERFECT DAY TO MAKE A KILLING

THE FEAR INDEX

ROBERT HARRIS
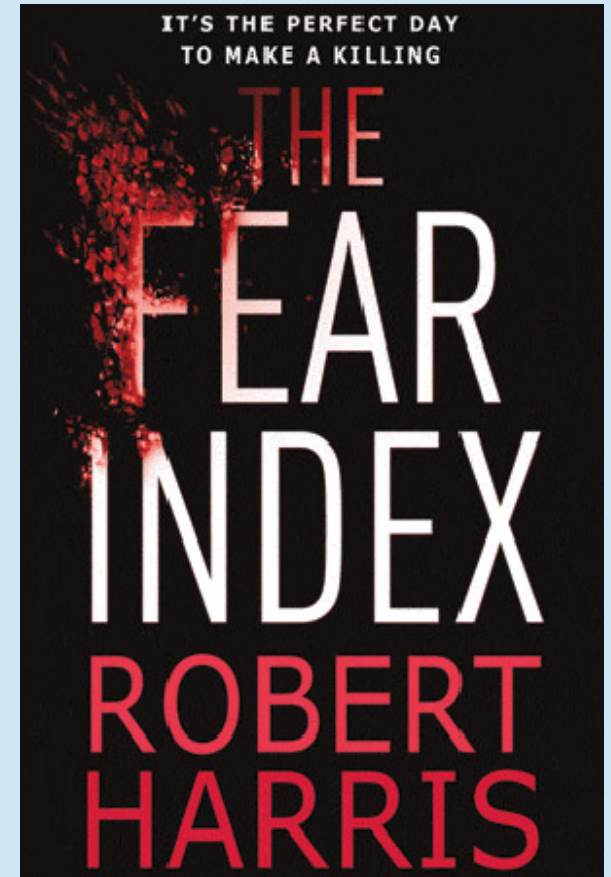
The Social Oscars 2013

Brandwatch ran a project to investigate how closely public opinion predicted/mirrored the results of the Oscars

What does big data have to do with social media analysis?

"Big Data is data that breaks Excel."

Or so the joke goes. But it's come to mean a lot more than that....

"Big data is an umbrella term that means a lot of different things, but to me, it means the possibility of doing extraordinary things using modern machine learning techniques on digital data. Whether it is predicting illness, the weather, the spread of infectious diseases, or what you will buy next, it offers a world of possibilities for improving people's lives."
- Shashi Upadhyay, CEO and Founder, Lattice Engines

"Big data is just the ability to gather information and query it in such a way that we are able to learn things about the world that were previously inaccessible to us."
- Hilary Mason, Founder of Fast Forward Labs

"Big data, which started as a technological innovation in distributed computing, is now a cultural movement by which we continue to discover how humanity interacts with the world — and each other — at large-scale."
 - Drew Conway, Head of Data, Project Florida

"Big data refers to the approach to data of "collect now, sort out later"…meaning you capture and store data on a very large volume of actions and transactions of different types, on a continuous basis, in order to make sense of it later. The low cost of storage and better methods of analysis mean that you generally don't need to have a specific purpose for the data in mind before you collect it.
 - Rohan Deuskar, CEO and Co-Founder, Stylitics

"Endless possibilities or cradle-to-grave shackles, depending upon the political, ethical, and legal choices we make."
 - Deirdre Mulligan, Associate Professor, UC Berkeley School of Information

Big Data is not new!

- Click to add Text

Staff sorting 4M used tickets from #London Underground
to analyse line use in 1939

The rise of Big Data

- "Big Data" was added to the OED in 2013.
- According to the Gartner Hype Cycle, it has now passed the "peak of inflated expectations" and is moving into the "trough of disillusionment".
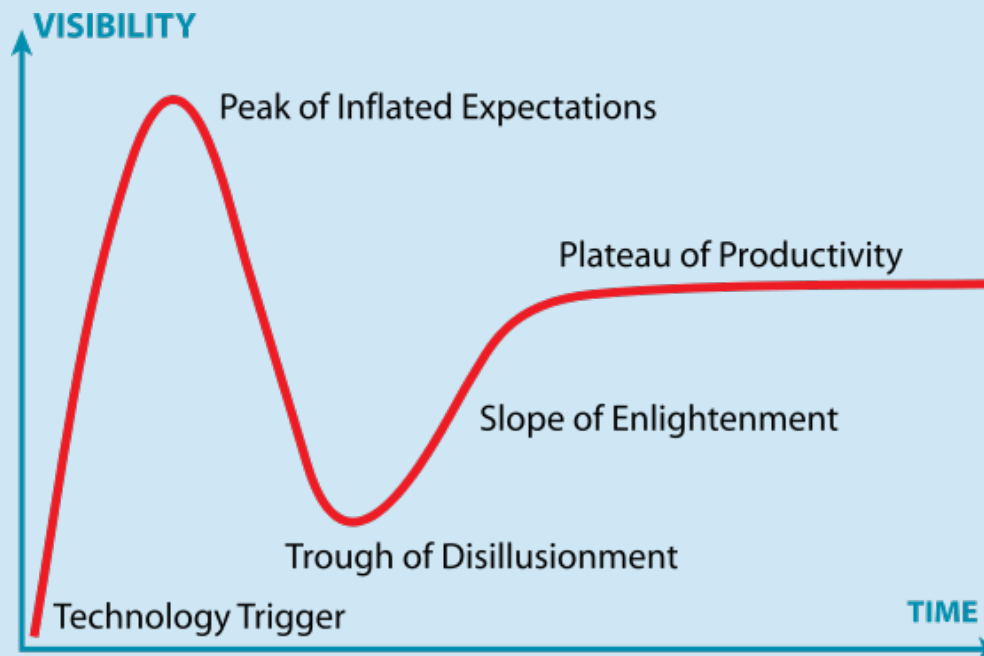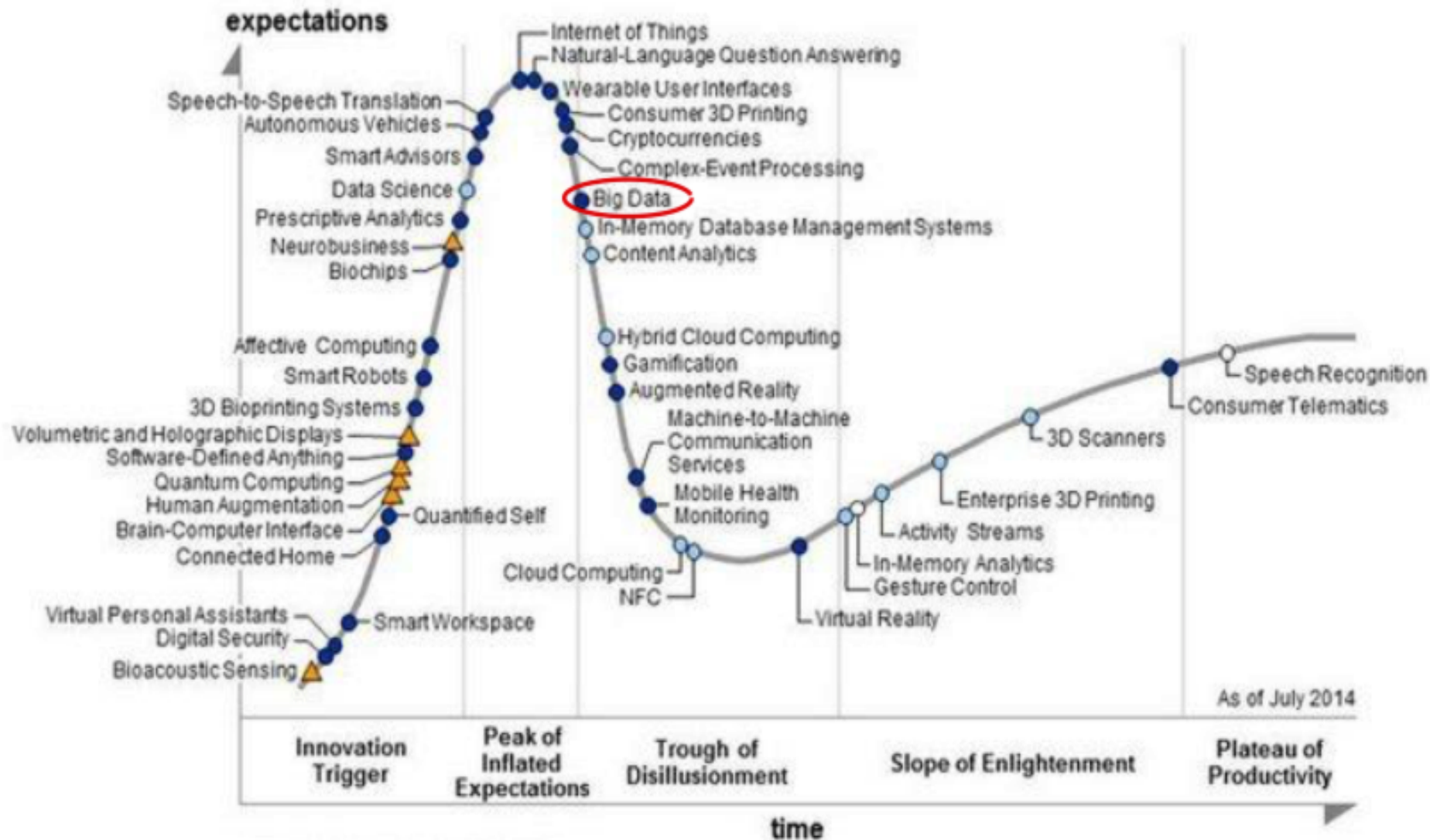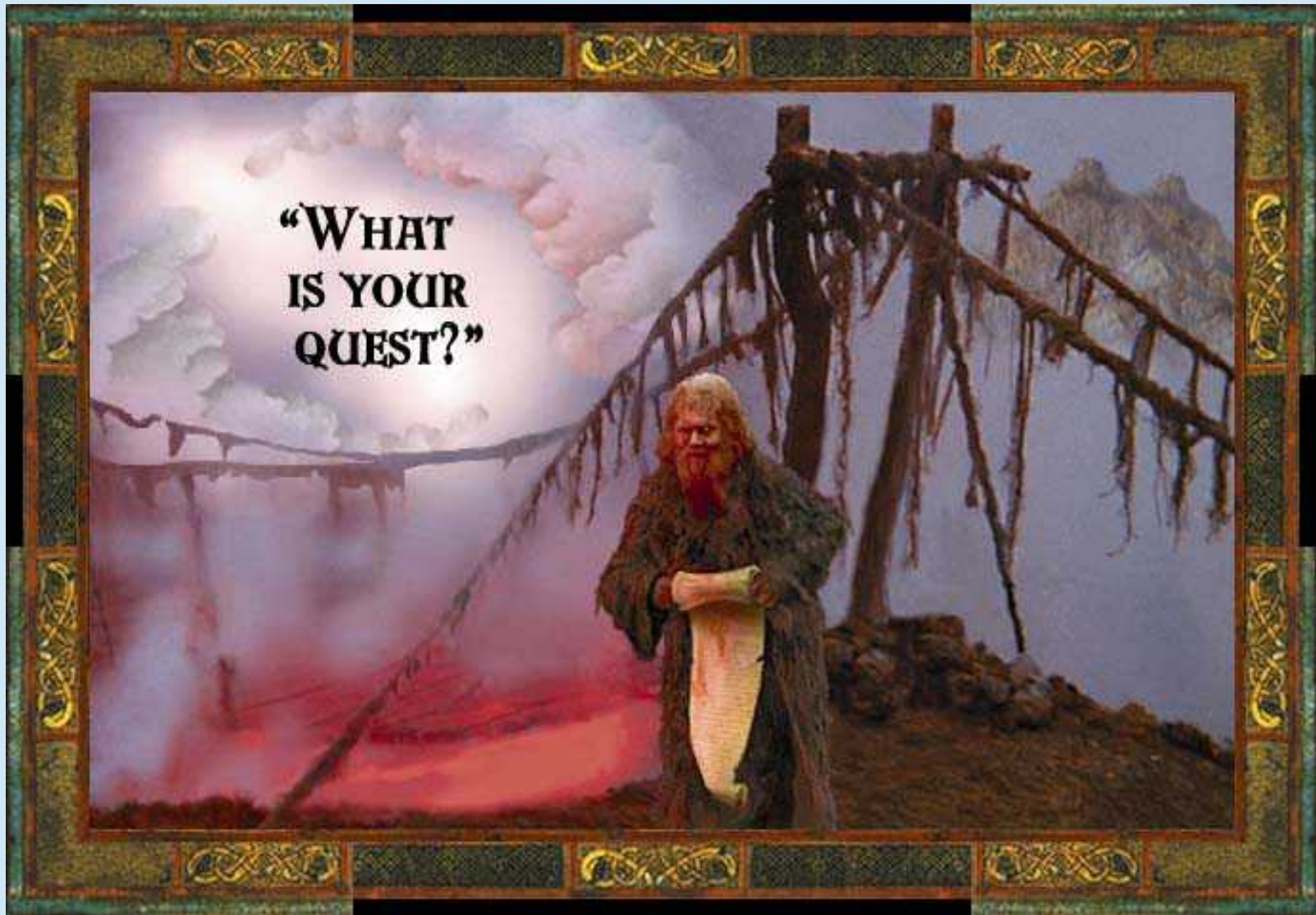
Figure 1. Hype Cycle for Emerging Technologies, 2014

Big Data:
the Holy Grail or the Bridge of Death?

How to Navigate the Bridge of Death

- A search for the Big Data Holy Grail can easily lead you to the "Bridge of Big Data Death"
- What business problem are you trying to solve with the help of Big Data?
- Will you fall into the Big Data Trough of Disillusionment?

Or are you King Arthur?
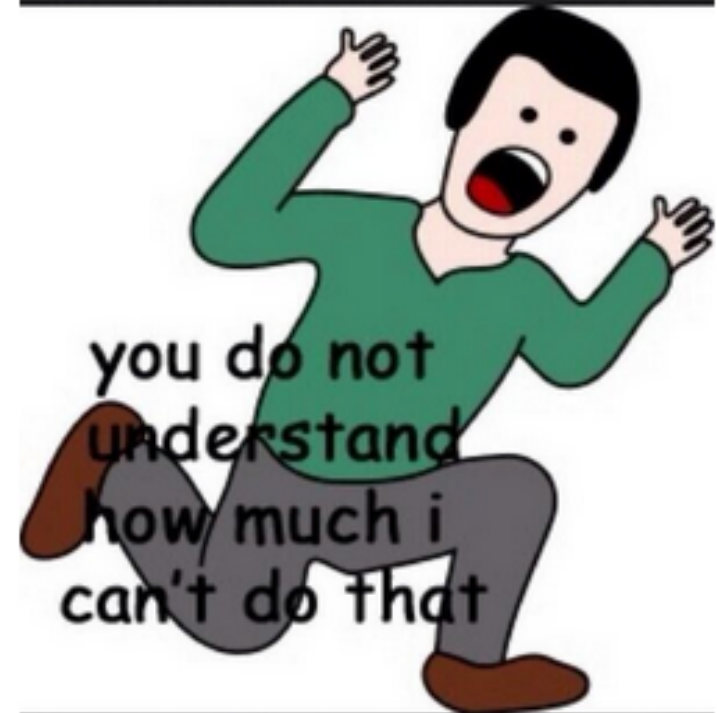
Big Data and Social Media

- Social media is not just about teenagers twittering about their favourite pop stars, or what they had for breakfast
- Although even this information can be useful, if you manufacture breakfast cereal or manage boy bands

Social Media is a valuable source of information

- business insights
- sharing and receiving important news
- campaigns
- all kinds of collective intelligence
- an alternative to traditional polls
- and much more

- One in six people have used social media to get information about an emergency
- 75% of people would use Facebook to post eyewitness information on an emergency or newsworthy event; 22% would use blogs, 21% would use Twitter
- During an emergency, one in two people would use social media to let loved ones know they are safe
- We need good real-time analysis tools to help process this data

Eoghan Rice @rice_e
This man survived by clinging to the roof of a house. Waves dragged him up to roof level. #yolanda #haiyan
pic.twitter.com/GsYu9JTq7S

📷 Expand   ← Reply   ↻ Retweet   ★ Favorite   ••• M

# Uses of social media during disasters

- Broadcasting info about the disaster
- Requesting info from local people and eyewitnesses
- Requesting and offering help and support
- Disaster mapping
- Mobilising the crowd to support initiatives

The clean-up in Clapham (Pic: Twitter via @Lawcol888)
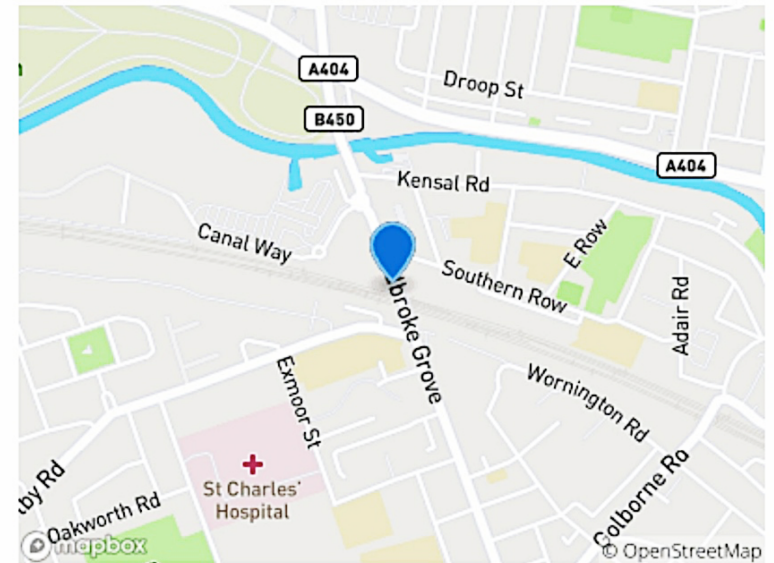
Harnessing the Crowd

- Using citizen reporters and digital responders for mapping crises

- Ushahidi deployed over 50k times
- Free and open source
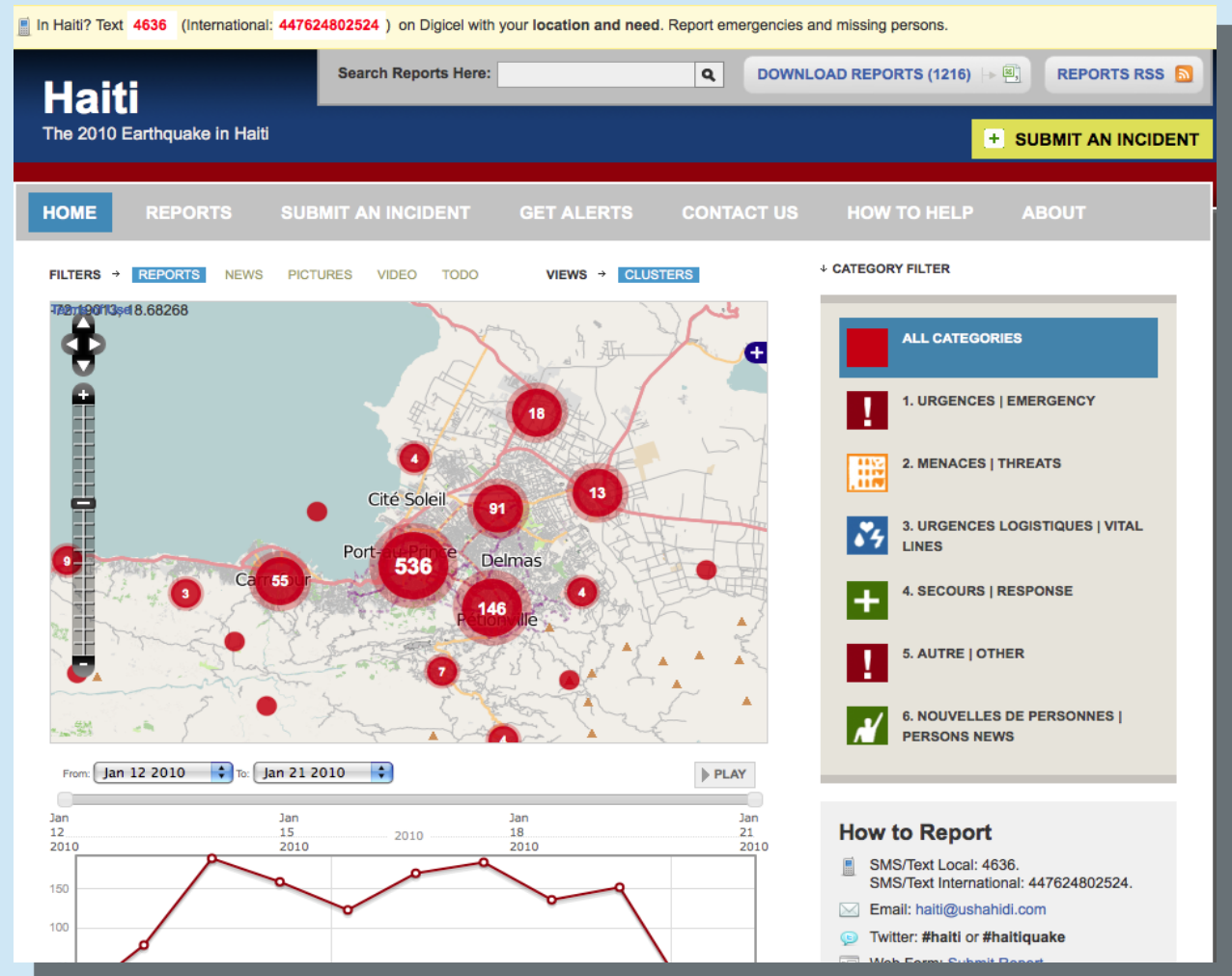
Tools to help disaster victims get aid quickly

Find mentions of locations in the text, match them to a knowledge base, and plot them on a map

Location

Annotation from YODIE

A huge fire has engulfed a tower block in West London. The fire broke out shortly before 1am on Wednesday at Grenfell Tower in Latimer Road near Notting Hill.The tower is at least 24 storeys high and contains 120 apartments. 200 firefighters have been tackling the blaze with 40 engines. A number of people have been treated for a range of injuries according to the fire brigade There have been multiple reports of people trapped in the blaze. These have not been confirmed by police or the fire brigade. Streets around the tower have been sealed off and residents in ther houses evacuated. London

30 injured in tube train crash

- How important and urgent is the message?
- What actions need to be taken?

Auto categories

🏷 related, derailment, affected_individuals

More than 30 people were taken to hospital after a crowded Tube train crashed into a tunnel wall, then hit the station platform, when it derailed in central London. Shocked passengers who were on board the Central Line train spoke of flying glass, and panic as they discovered some carriage doors were jammed shut. No-one was seriously injured. Fleets of ambulances surrounded the exit to the station as a major emergency services operation swung into action following the incident, which occurred at around 1.50pm on Saturday. A London Underground spokeswoman said it was believed all 800 passengers had been taken off the train and casualties had been taken to three nearby hospitals, but the Fire Brigade said crews were still searching the scene.

DisasterAWARE
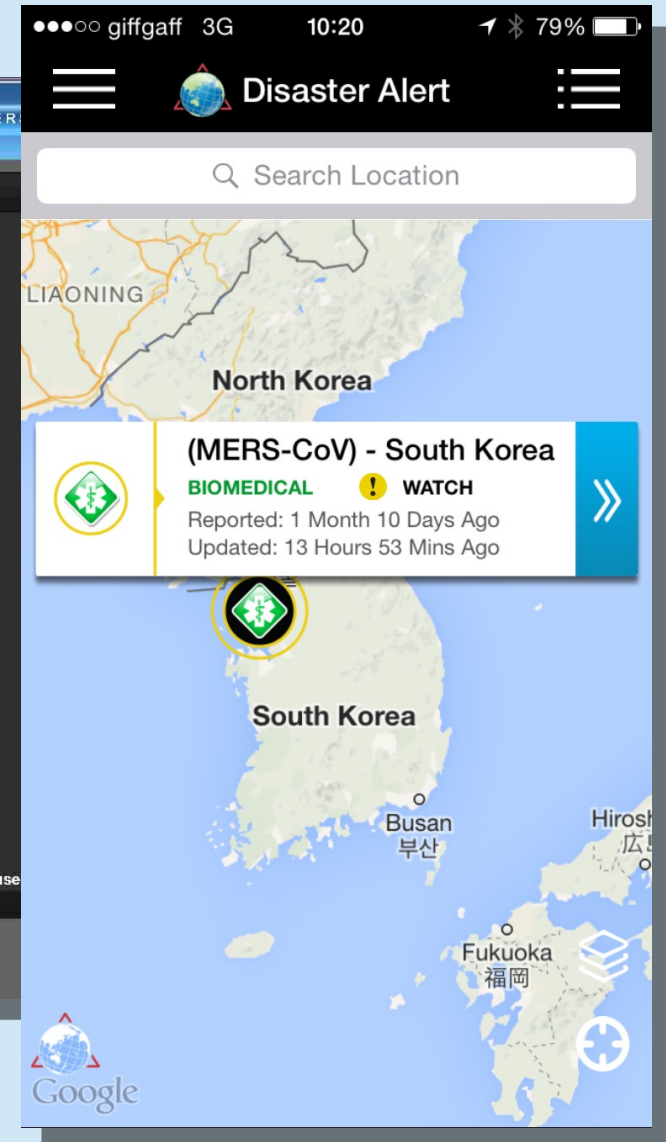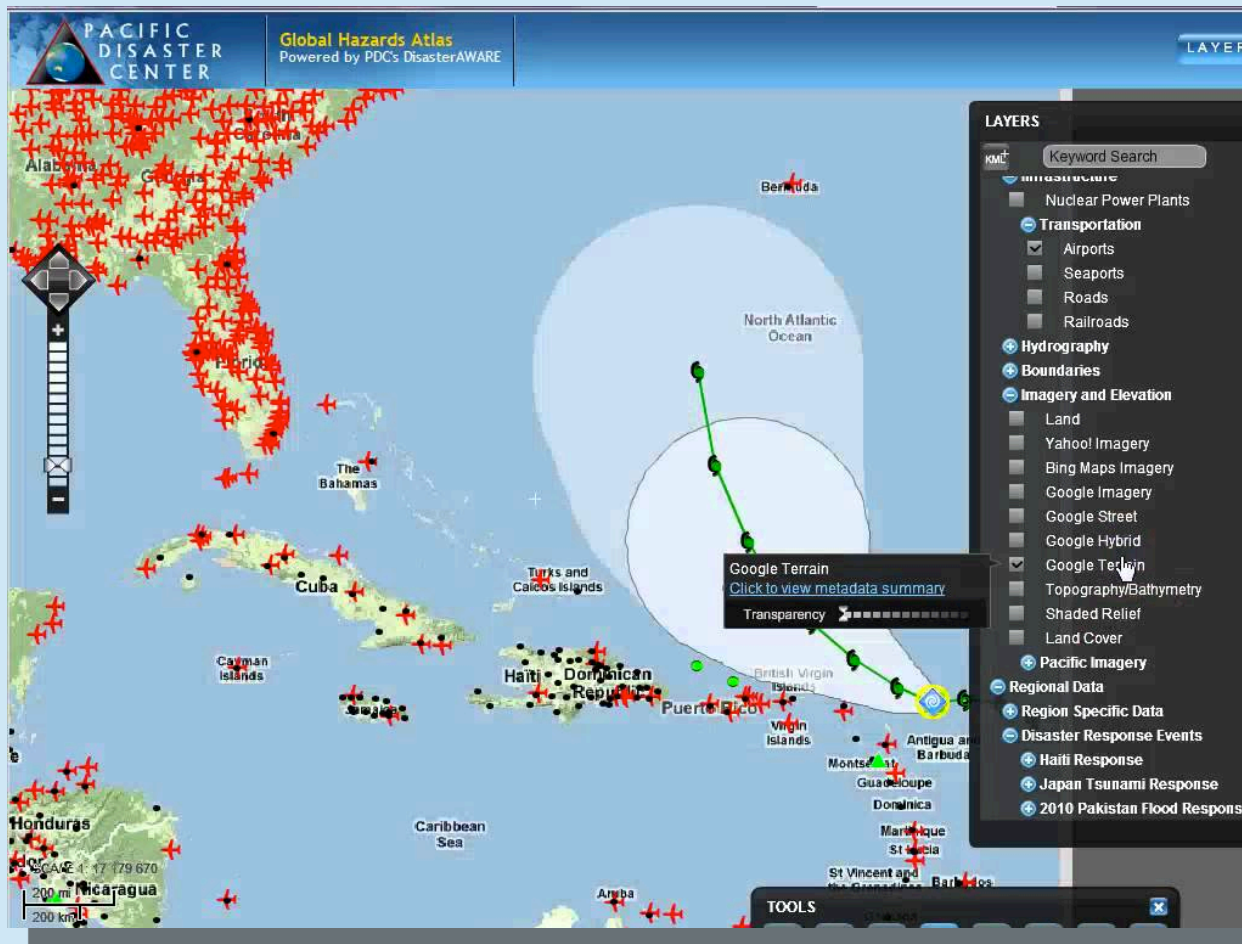
Early warning and decision support platform

NLP tools give us a way to understand the data

- **sort** the data to remove the drivel from the interesting parts
- **extract** the relevant pieces of information
- **link** the extracted information to other sources of information (e.g. from DBpedia)
- **aggregate** the information according to potential new categories
- **query** the (aggregated) information
- **visualise** the results of the query

Where in the UK did MPs tweet more about the economy?

SCOTLAND
17.15% of the tweets about UK Economy were made by MPs from this region.
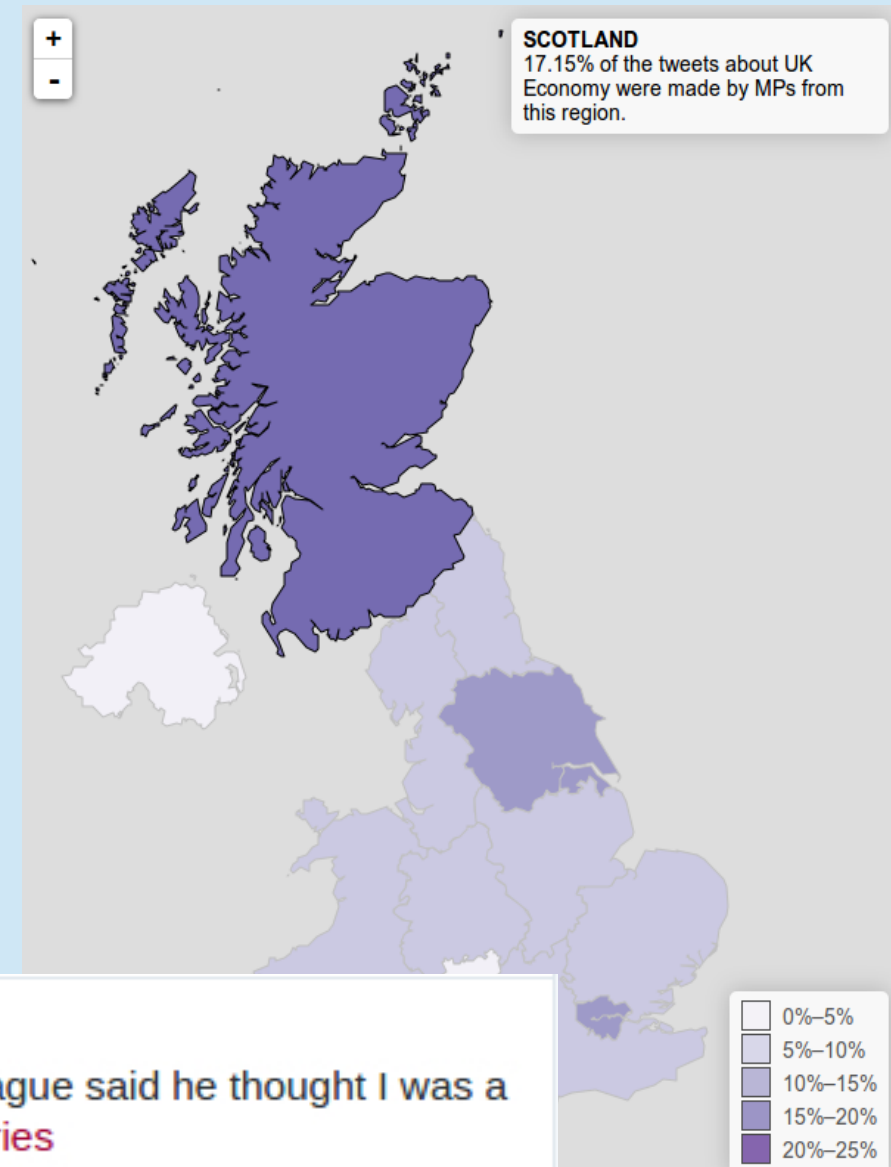
Richard Short @ToryShorty · May 28
I spoke up for the living wage today and a work colleague said he thought I was a Tory. He obviously hasn't heard of #WorkingClassTories

↩   ⇄ 2   ★ 5   +👤   •••

0%–5%
5%–10%
10%–15%
15%–20%
20%–25%

Leaflet | Statistical Data © Nesta, 2014-2015

Find all tweets where a Conservative MP talked about the economy

Searching Index "2015-03-09"

{DocumentAuthor author_party="Conservative Party"} OVER {Topic theme="uk_economy"}

Search

Richard Short
@ToryShorty

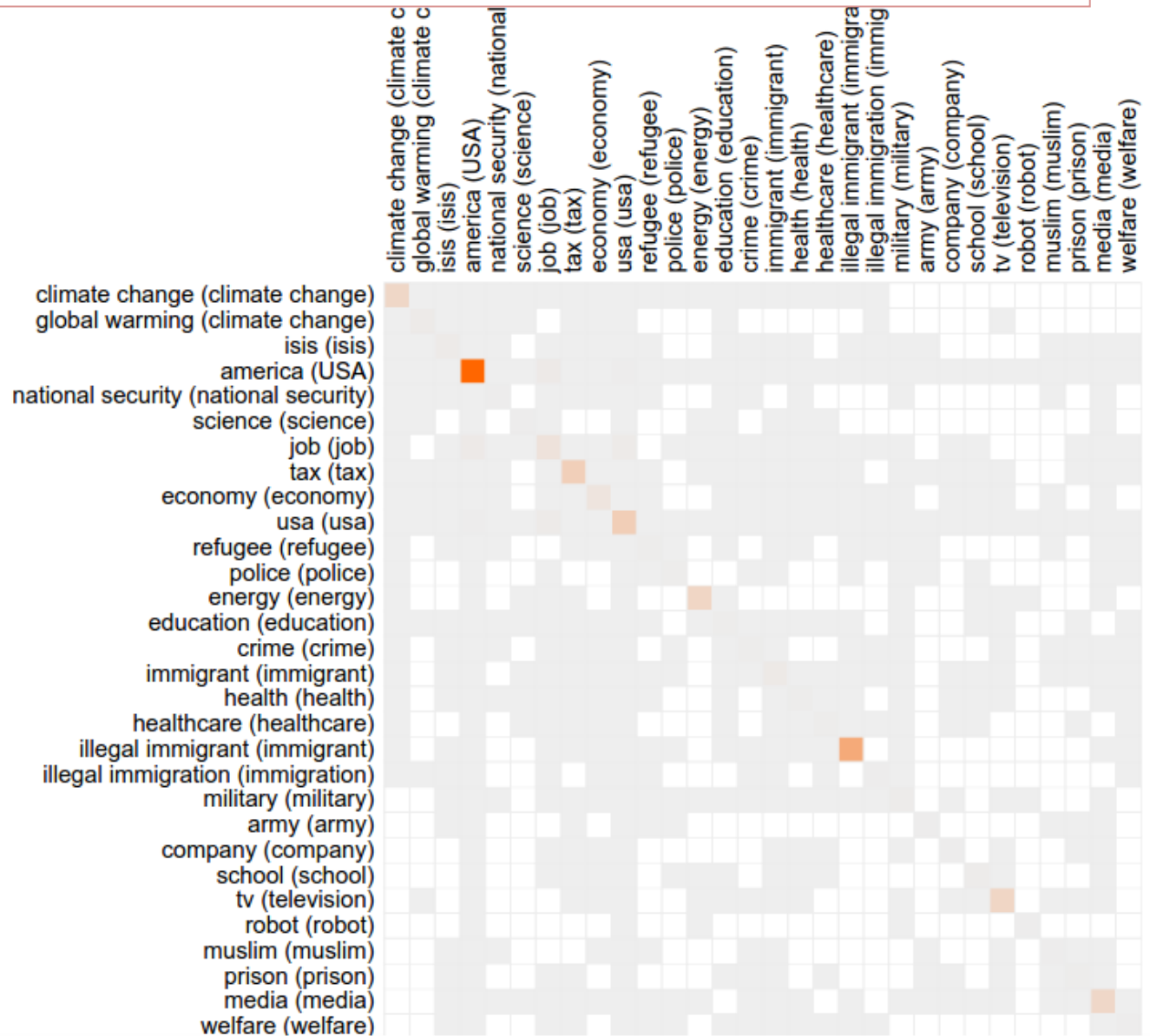With so many protected from Labour's pension raid are they sure it will even generate £2.7bn #bbcsp
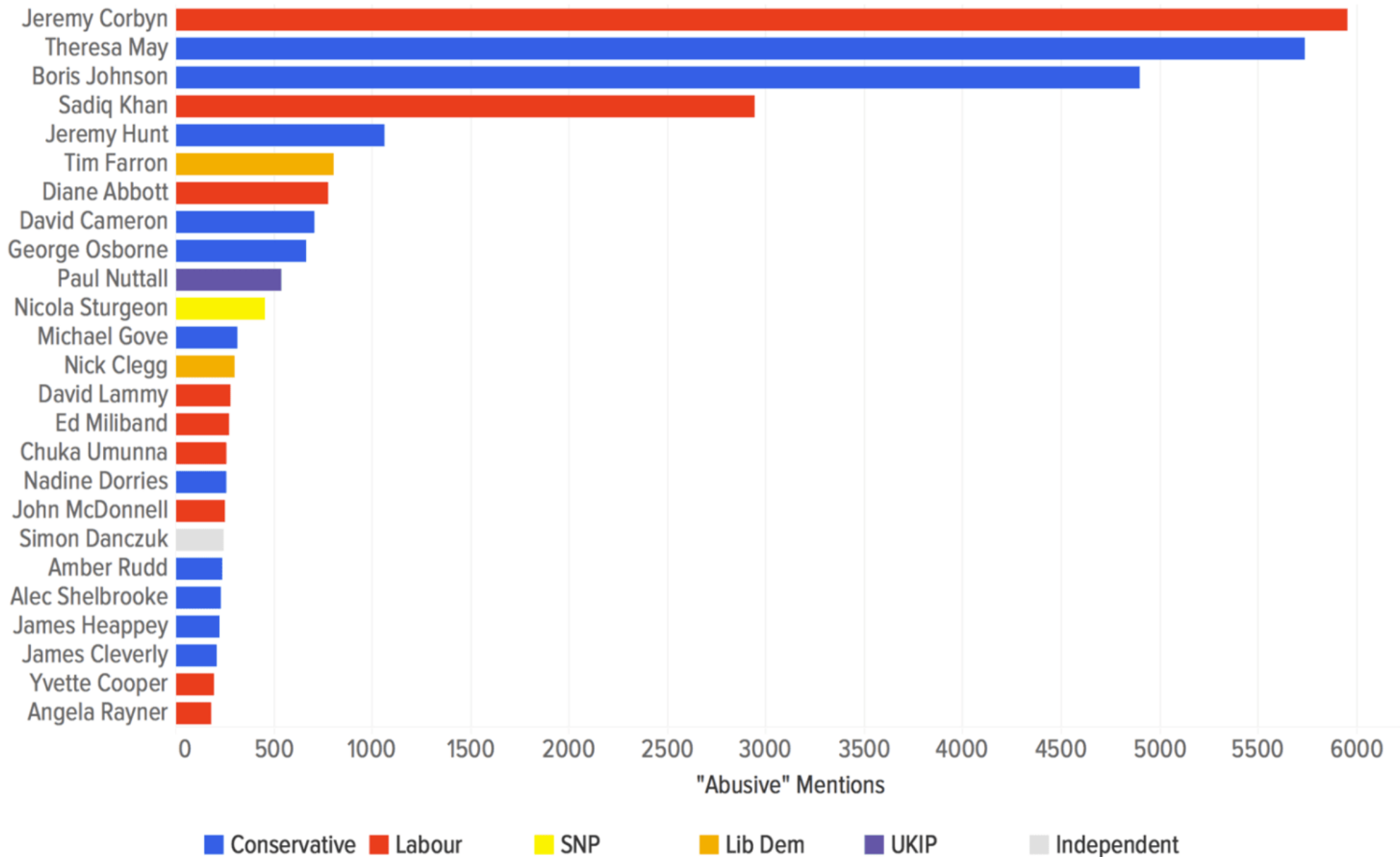
What is Trump ranting about?

Good luck with this one!

33

Which political group gets the most hate speech?
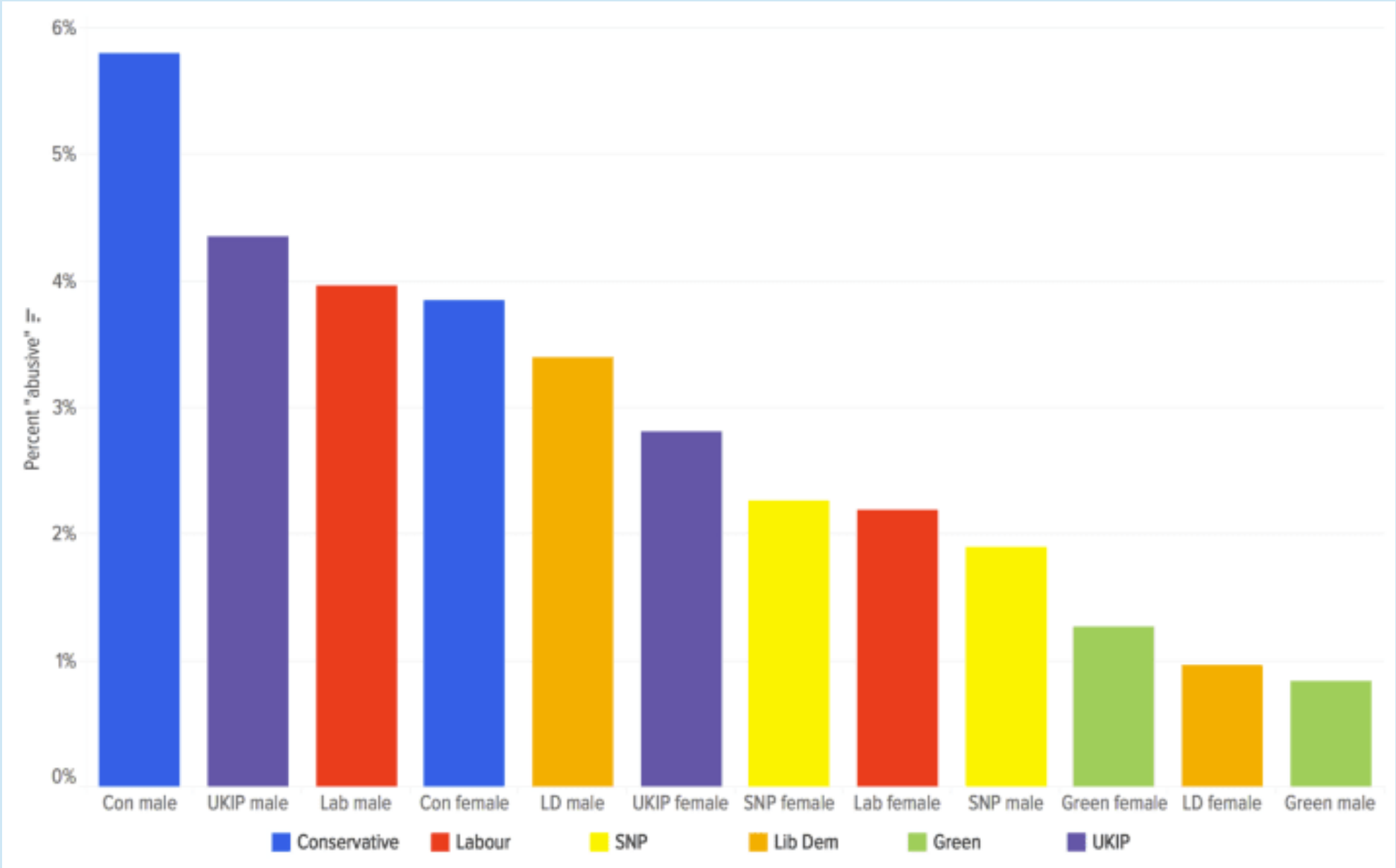
Which MP?

What words did they use for men?

Percent of total male abuse

Word	
fuck off	~9.8%
idiot	~8.8%
cunt	~8.0%
kill	~6.4%
twat	~5.6%
scum	~4.0%
prick	~3.3%
coward	~3.2%
moron	~2.7%
die	~2.2%

What about women?

Percent of total female abuse
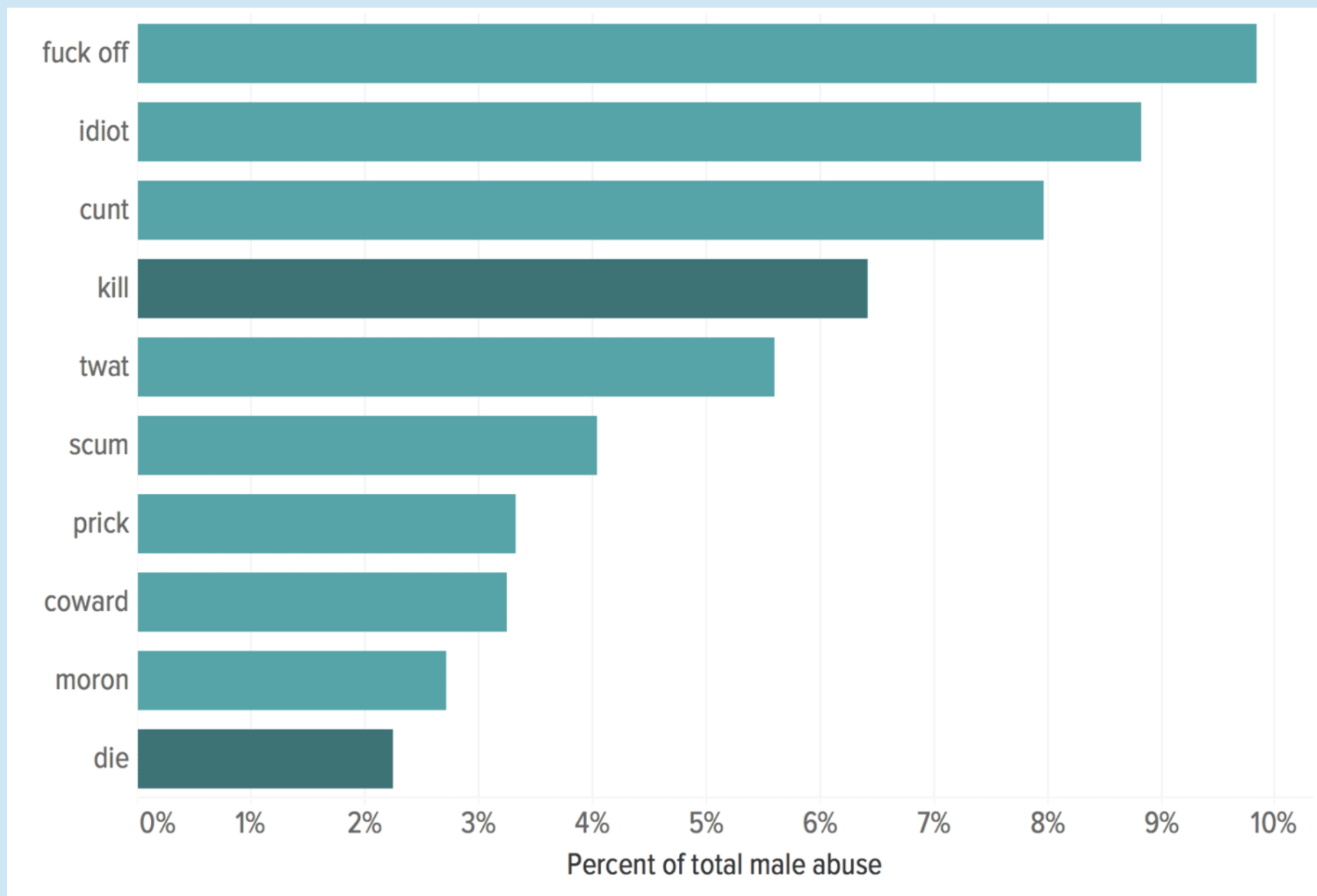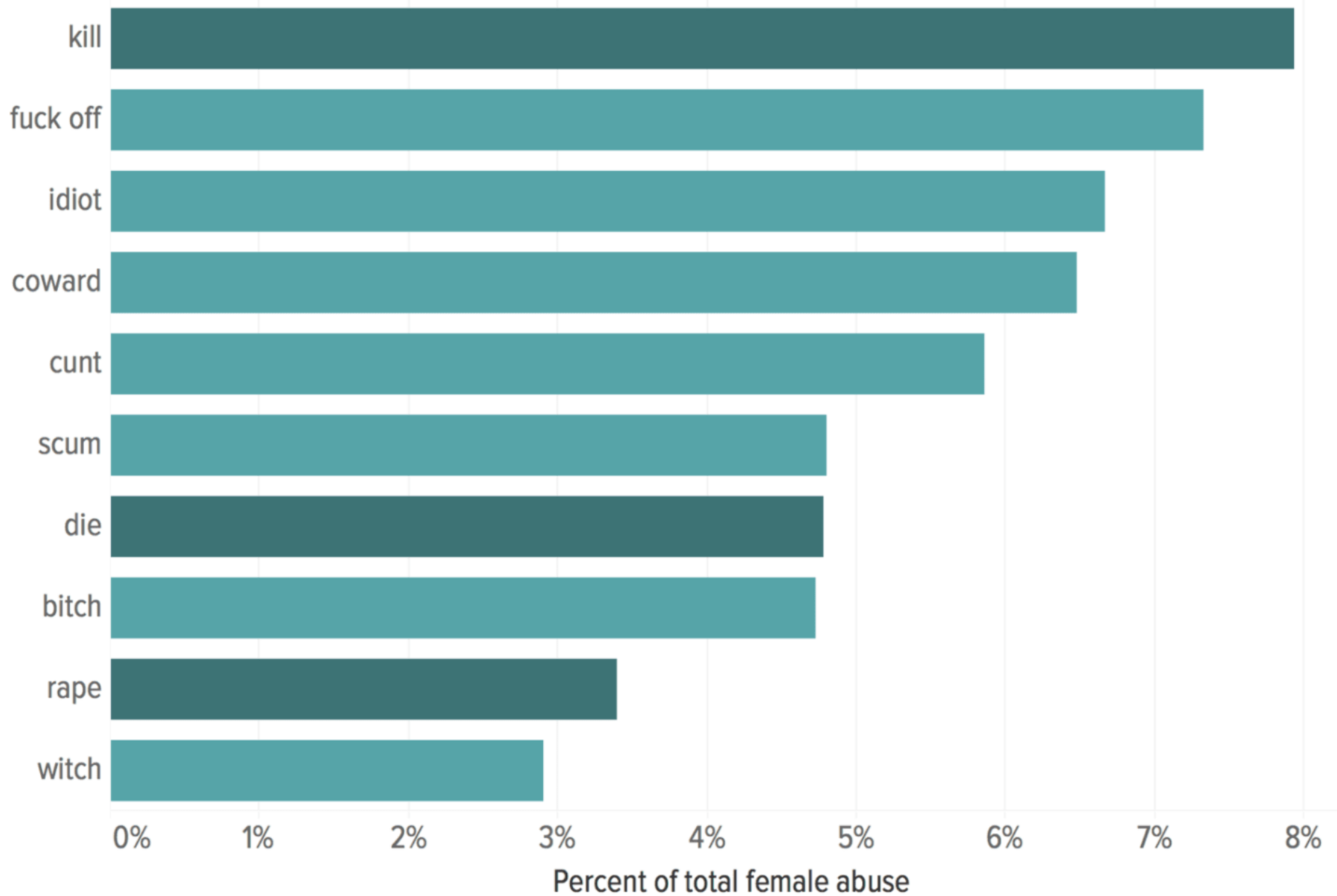
Left-wing politicians got more abuse after terrorist attacks
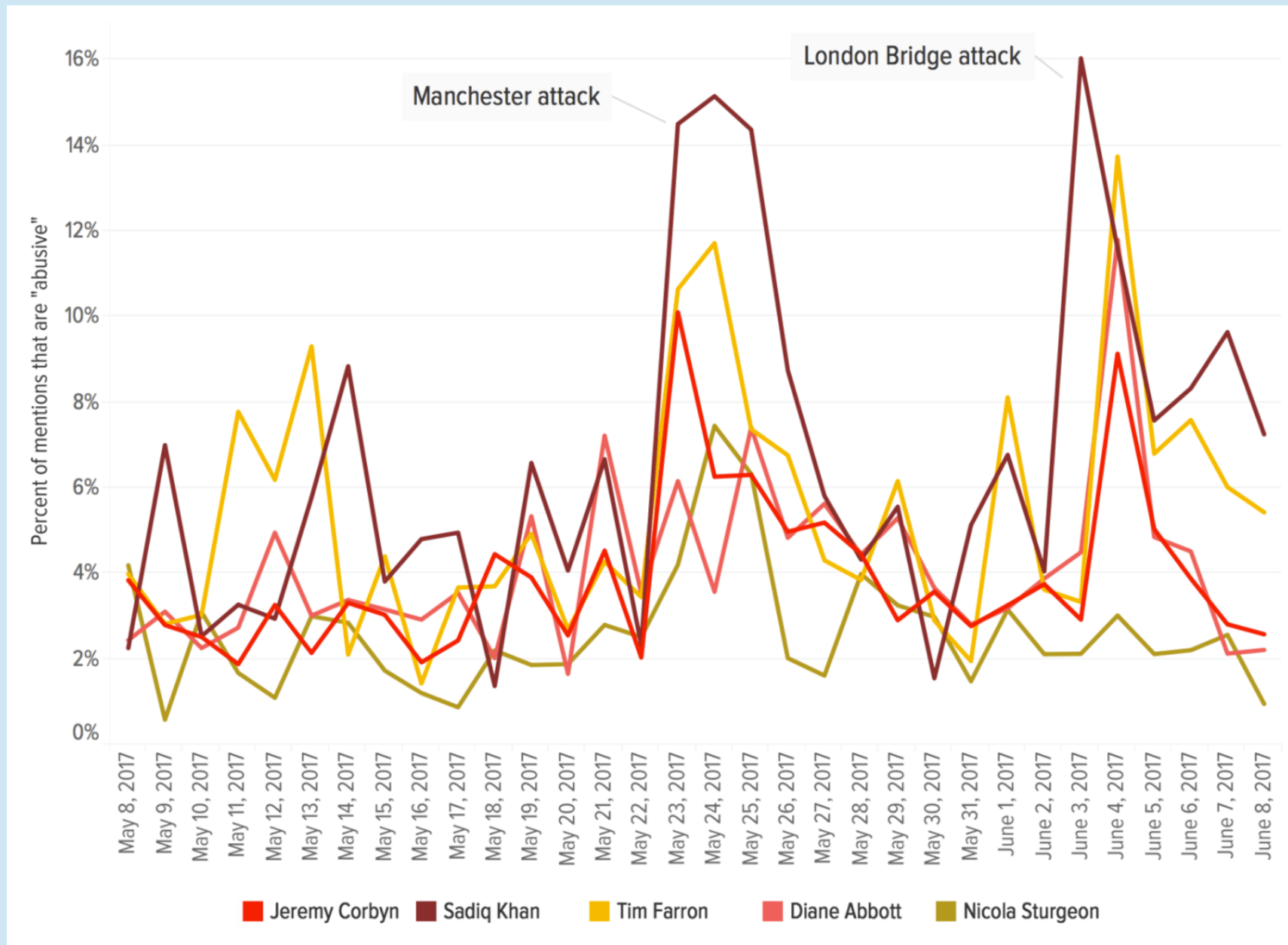
Summary

- This first part has just given a very brief overview of what the point of social media analysis is
- In the rest of the course, we'll look at all these things in more detail, including problems and solutions