

# **Computing + Statistics = Data Science**

Dr Graham Cormode









**You** **Tube**





L'ORÉAL



Coca-Cola





# Data Science:

“What can we do with  
all this data?”

10 minutes of data analysis

## Data Science Questionnaire

### Demographics

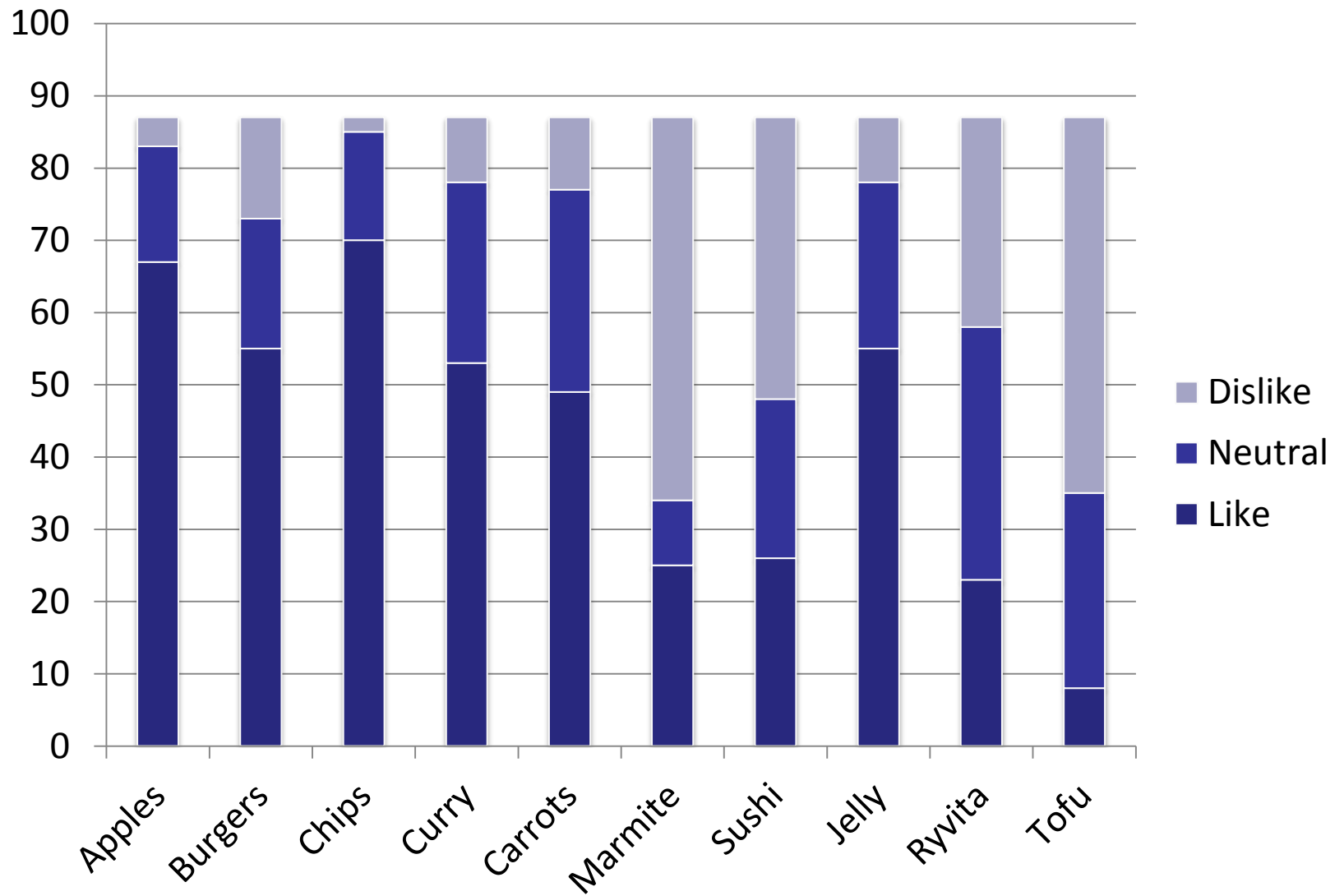
Female  Male

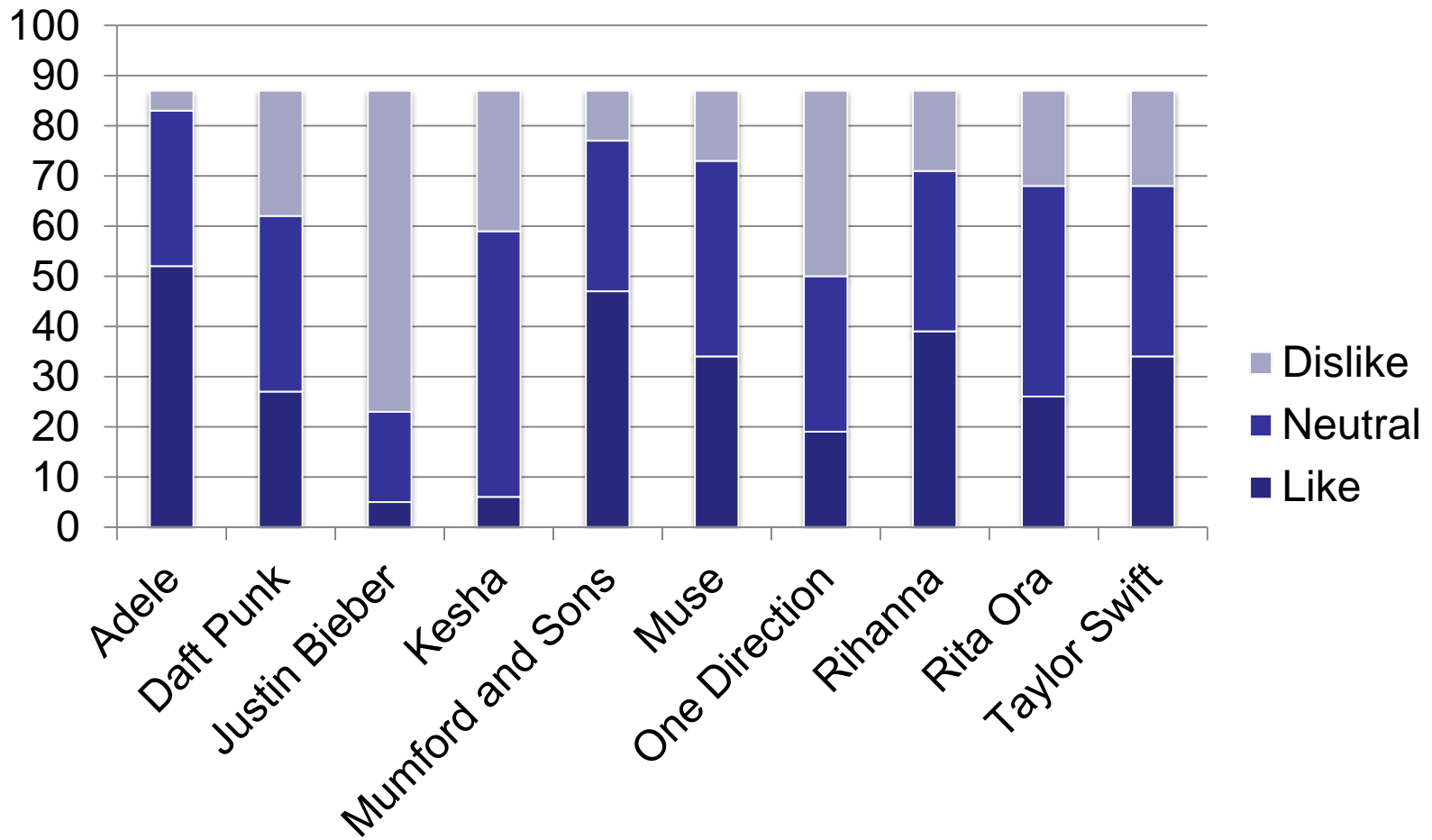
### Music

	Like	Neutral	Dislike
Adele	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Daft Punk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Justin Bieber	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ke\$ha	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mumford & Sons	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Muse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
One Direction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rihanna	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rita Ora	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Taylor Swift	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

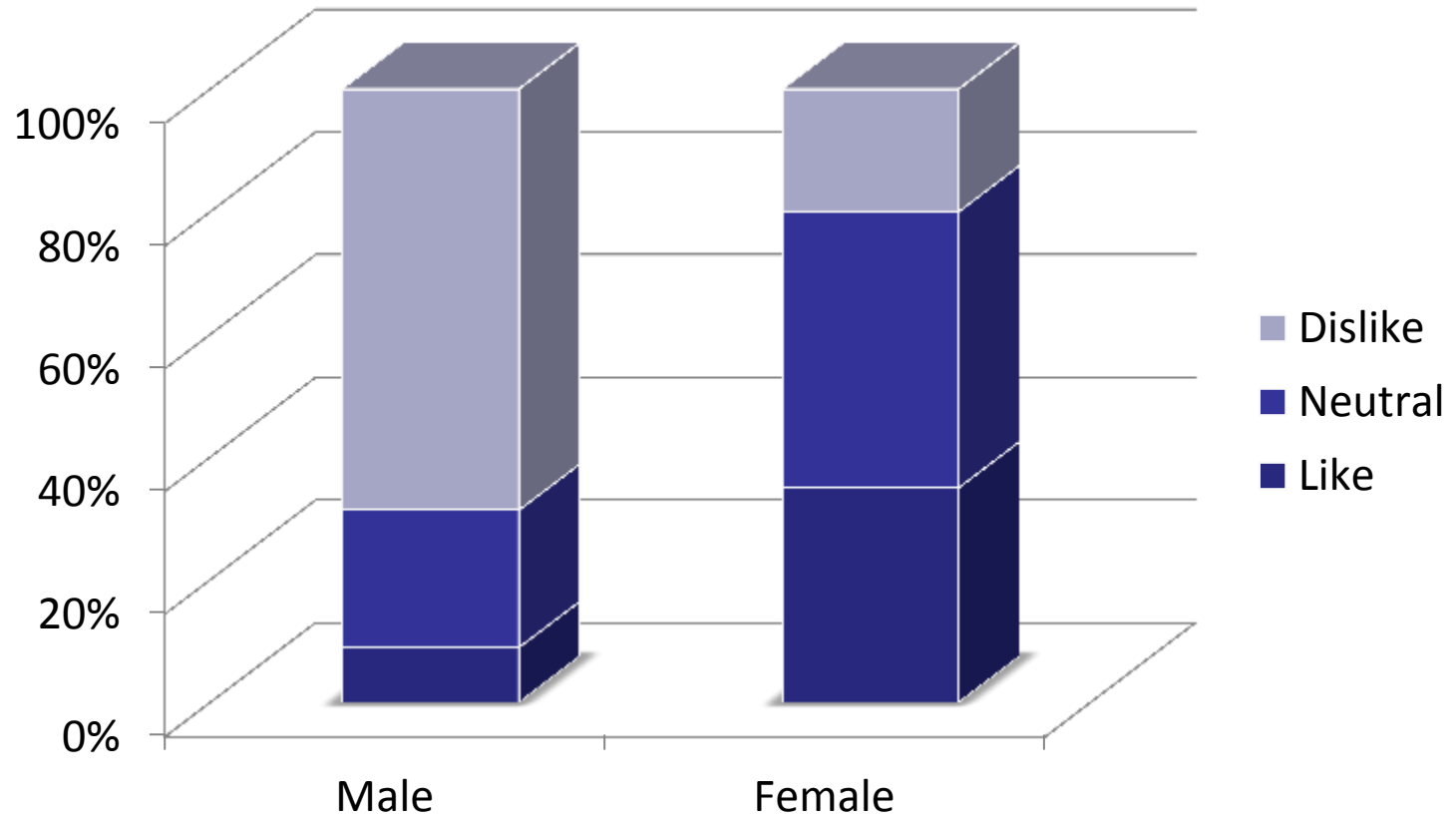
### Food

	Like	Neutral	Dislike
Apples	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Burgers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chips	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Curry	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Carrots	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Marmite	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Baked beans	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chocolate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ryvita	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tofu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>





# Ke\$ha



Like Rihanna --> Like Adele (81%)

Like Muse --> Dislike Justin Bieber (88%)

Dislike One Direction --> Dislike Justin Bieber (97%)

Dislike One Direction --> Like Muse (56%)

Like Burgers --> Like Chips (92%)

Like Ryvita --> Like Apples (96%)

Like Marmite --> Like Curry (80%)

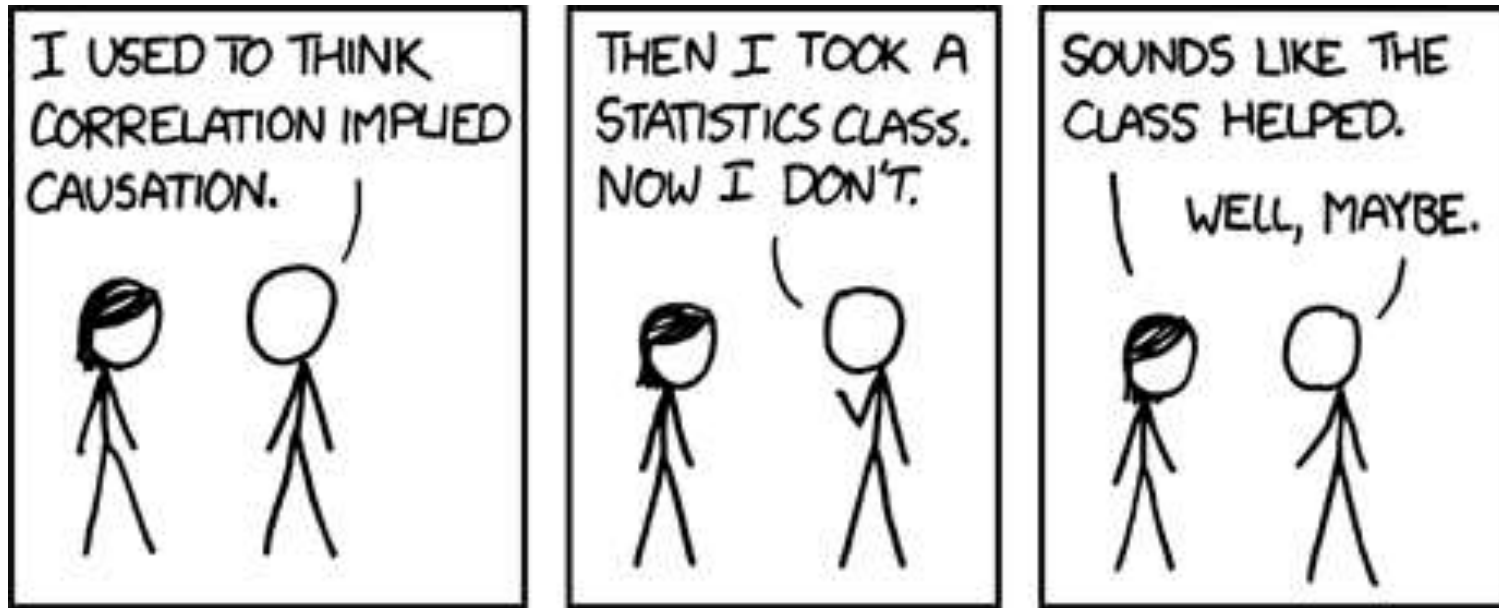
Dislike Sushi --> Dislike Marmite (71%)



Like Muse --> Like Curry (79%)

Like Rihanna --> Dislike Marmite (68%)

Like Taylor.Swift --> Dislike Tofu (64%)



<http://xkcd.com/552/>

## Your Amazon.co.uk

**Featured Recommendations**

[Books](#)

[Health & Personal Care](#)

[Electronics](#)

[Computers & Accessories](#)

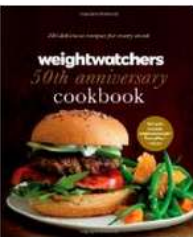
[Toys & Games](#)

[Jewellery](#)

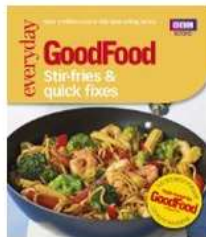
[See All Recommendations](#)

### Books

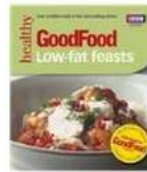
Page 1 of 17



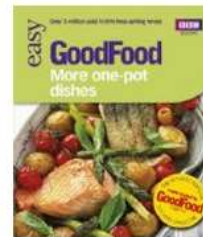
**New Release**  
Weight Watchers 50th...  
Weight Watchers  
Hardcover  
£19.69 £17.31  
[Why recommended?](#)



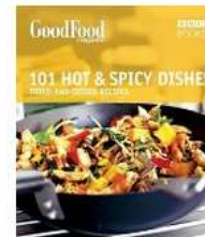
**New Release**  
Good Food  
Bbc Books  
Paperback  
£4.99 £3.74  
[Why recommended?](#)



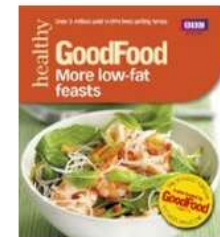
Good Food  
Good Homes Magazine  
★★★★☆ (108)  
Paperback  
£4.99 £3.74  
[Why recommended?](#)



Good Food  
Jane Hornby  
★★★★☆ (18)  
Paperback  
£4.99 £3.74  
[Why recommended?](#)



Good Food  
Orlando Murrin  
★★★★☆ (12)  
Paperback  
£4.99 £3.74  
[Why recommended?](#)



Good Food  
Sharon Brown  
★★★★☆ (29)  
Paperback  
£4.99 £3.74  
[Why recommended?](#)



[See all recommendations in Books](#)



## Discover more music

Last.fm is a music discovery service that gives you personalised recommendations based on the music you listen to.

[Start your profile](#)**Muse**

3,531,467 listeners

alternative rock · rock

**The Beatles**

3,064,822 listeners

classic rock · rock

**Calvin Harris**

1,264,151 listeners

electronic · dance

[Google.org home](#)

[Dengue Trends](#)

Flu Trends

Home

Select country/region

[How does this work?](#)

[FAQ](#)

Flu activity



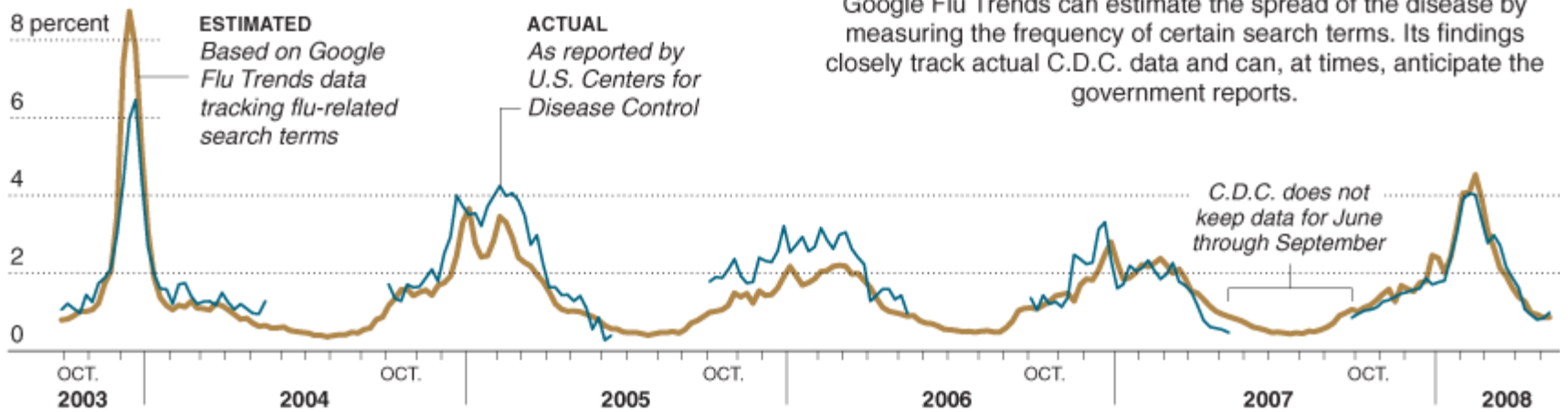
### Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)



[Download world flu activity data](#) - [Animated flu trends for Google Earth](#) - [Compare flu trends across regions in Public Data Explorer](#)

### PERCENT OF HEALTH VISITS FOR FLU-LIKE SYMPTOMS *Mid-Atlantic region*



Sources: Google; Centers for Disease Control

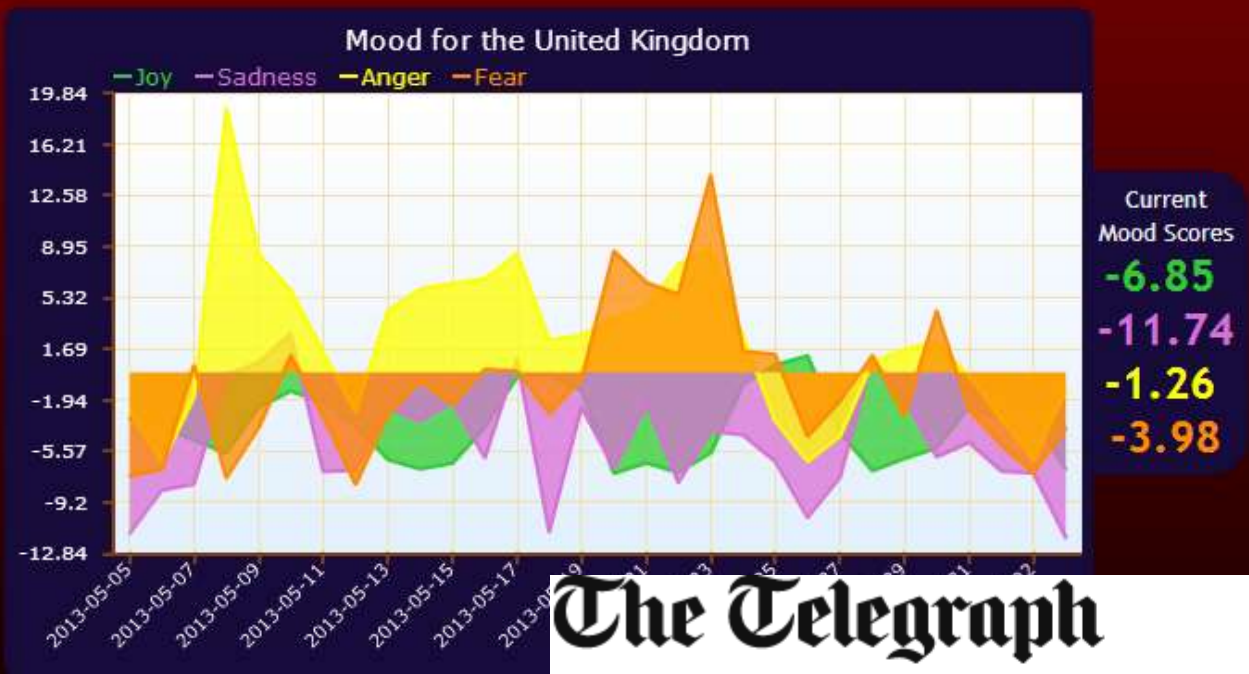
### Using Google to Monitor the Flu

Google Flu Trends can estimate the spread of the disease by measuring the frequency of certain search terms. Its findings closely track actual C.D.C. data and can, at times, anticipate the government reports.



Based on geolocated Twitter content, we track four types of Mood and Affect (Joy, Sadness, Anger and Fear). This figure shows the daily mood time series for the entire United Kingdom in the last 31 days

Check Regional Mood Scores:  
for the last 31 days  
since July 2009



Mood of the Nation uses the content of Twitter to nowcast mood rates in several UK regions. This page is still in BETA. Parts of the methodology together with relevant findings are described in the following publications:

*Detecting Events and Patterns in Large-Scale User Generated Textual Streams with Statistical Learning Methods*, V. Lampos, Ph.D. Thesis, 2012.

*Effects of the Recession on Public Mood in the UK*, T. Lansdall-Welfare, V. Lampos and N. Cristianini, MSND/SMANE at WWW 2012.

Check also the [Twitter Mood animation](#)

Vertical axis: Normalised mood scores for each category positive value negative values show the opp  
Horizontal axis: Date in yyyy-mm-dd format

## The Telegraph

HOME NEWS WORLD SPORT FINANCE COMMENT BLOGS CULTURE TRAVEL LIFE  
Technology News | Technology Companies | Technology Reviews | Video Games | Technology Video

HOME » TECHNOLOGY » TWITTER

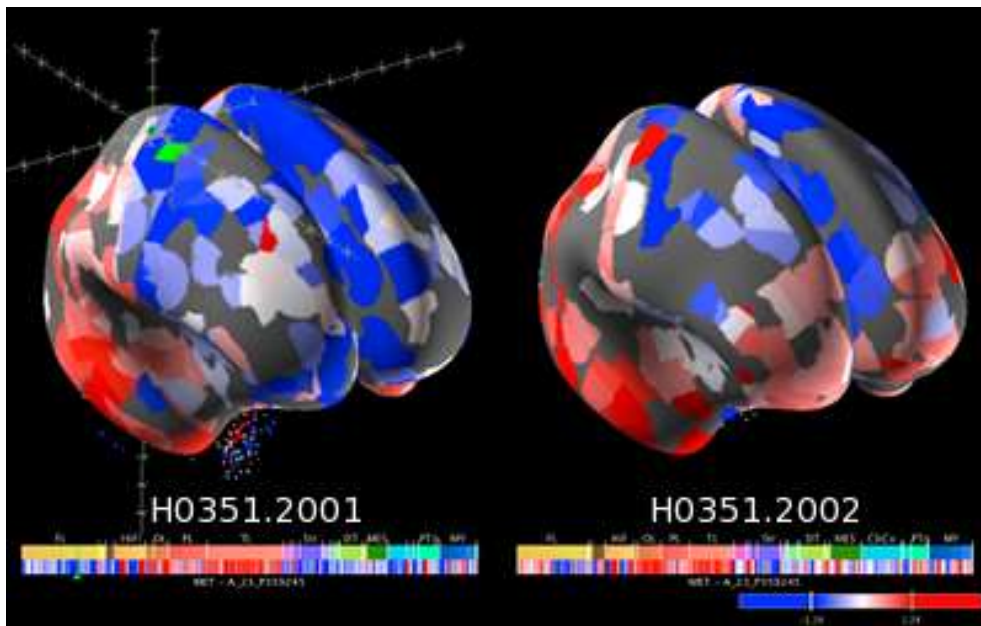
### Rise in public anger found on Twitter before riots

The summer riots that spread across the country from London last year could have been predicted by an analysis of Twitter, new research has suggested.

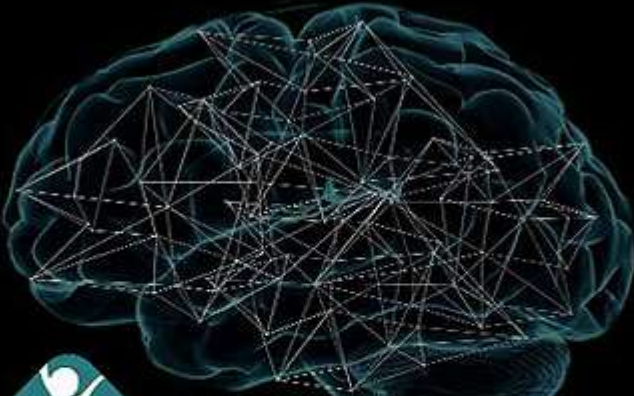
What are the big questions?









**ALLEN INSTITUTE**  
*for* **BRAIN SCIENCE**  
*Fueling Discovery*




**HBP**  
 The Human Brain Project  
[www.humanbrainproject.eu](http://www.humanbrainproject.eu)

**WARWICK**



# Babel Fish

Instant Translation



# NOTEBOOK

The notebook having horizontal ruled line and being able to fold up is the best for arranging sentences.







# CENTER FOR URBAN SCIENCE AND PROGRESS – NEW YORK

---

“OVER THE NEXT FIVE YEARS, 370 JAY STREET WILL BE TRANSFORMED INTO A CUTTING-EDGE CENTER FOR RESEARCH AND SCIENCE THAT WILL GIVE ANOTHER HUGE BOOST TO OUR CITY’S ECONOMY. OUR ADMINISTRATION HAS LONG SEEN THE PROMISE OF DOWNTOWN BROOKLYN, AND WE’VE MADE THE INVESTMENTS NEEDED TO TRANSFORM IT INTO A THRIVING CENTER FOR BUSINESS. WITH THE ADDITION OF THIS NEW CAMPUS, BROOKLYN WILL BE ONE OF THE MOST DYNAMIC ENVIRONMENTS FOR ENTREPRENEURS ANYWHERE IN THE COUNTRY.”

MAYOR MICHAEL R. BLOOMBERG

---

**CRIME SCENE-DO NOT ENTER**





# What are the tools?

- Foundations
  - Maths
  - Computer Science
  - Statistics
  - Sociology
- Concepts
  - Machine learning/AI
  - Data mining
  - Software Engineering: Big data



# Lots more information online

- A Very Short History Of Data Science

<http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/>

Data Scientist: The Sexiest Job of the 21st Century

<http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

The role of Statistics in the Higgs Boson discovery

<http://blog.revolutionanalytics.com/2012/07/discovering-the-higgs-boson-with-statistics.html>

Questions?