

Data,

**Bio
data
ata**

**& social
media data**

Susan Halford, Leslie Carr

Web Science Institute, University of Southampton



WEB SCIENCE
Institute

UNIVERSITY OF
Southampton

Part I: Computer Science Basis for Data Analytics

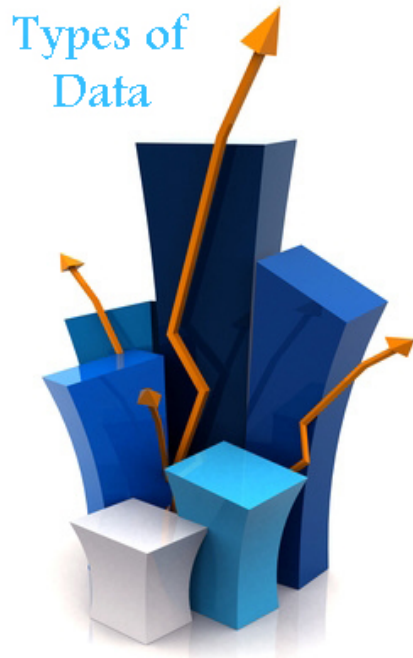
Leslie Carr

Algorithms + Data Structures = Programs

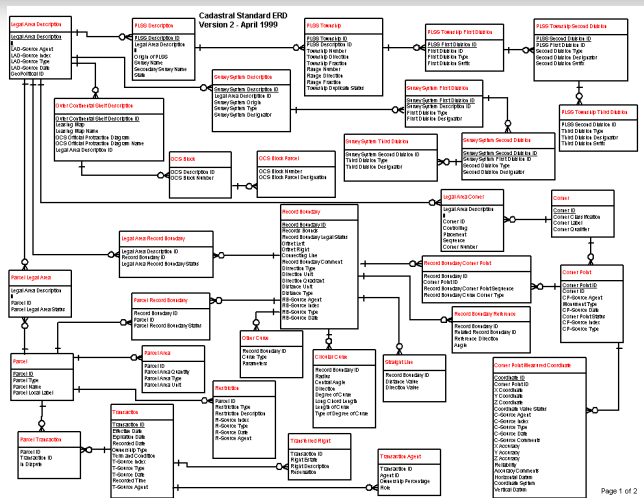
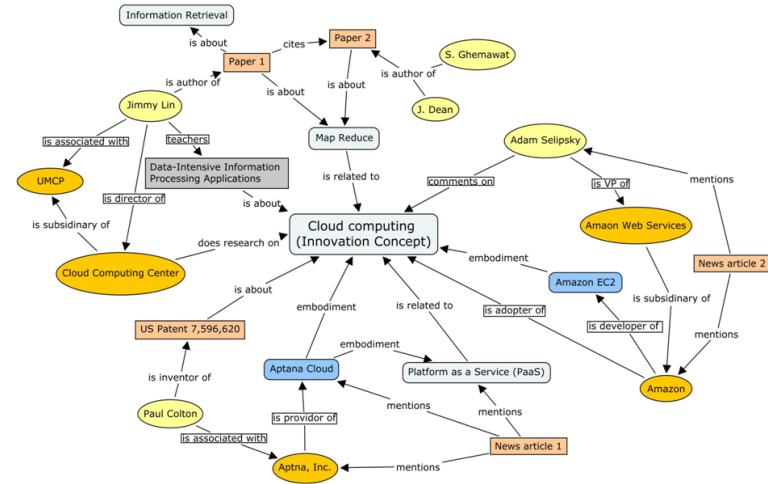
- Computer science is about
 - Finding solutions to problems
 - By modelling situations using computational techniques
 - Abstraction, generalisation, decomposition
 - Data structures
 - Algorithms
- Mainly assumes that someone defines the problem, someone uses the solution
 - Specification
 - Design
 - Use
- Corporate IT
- Domestic / personal devices

Data Handling

- We can reduce problems to repeating sets of isomorphic data
- Because someone controls the data or the problem



Data: Type, Schema



Students

ID#	Name	Phone	DOB
500	Matt	555-4141	06/03/70
501	Jenny	867-5309	3/15/81
502	Sean	876-9123	10/31/82

ID#	ClassID	Sem
500	1001	Fall02
501	1002	Fall02
501	1002	Spr03
502	1003	S203

Takes_Course

ClassID	Title	ClassNum
1001	Intro to Informatics	I101
1002	Data Mining	I400
1003	Internet and Society	I400

Courses

1980s: Make Your Problem Fit Inside This



1980s: Make Your Problem Fit Inside This



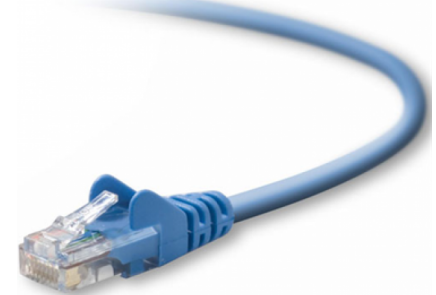
Extend it with
700Mb of prepared,
commercial data on
this



1990s: Use Data from Internet/Web 1



Extend it with Gb of prepared, commercial data on the Internet



2008- Use Data from Web 2



Extend it with Gb of
contributed information
from the Web



2008- Use Data from Web 2



A view outside

- IRL
- meatspace



Big Data

- Volume
 - 1bn websites, 0.5bn tweets/day, 16yrs of new YouTube videos/day
- Variety
 - text, image, video
- Velocity
 - flow
- Veracity
 - What is truth
- Value
 - new kinds of data = new kinds of science = new kinds of economic value

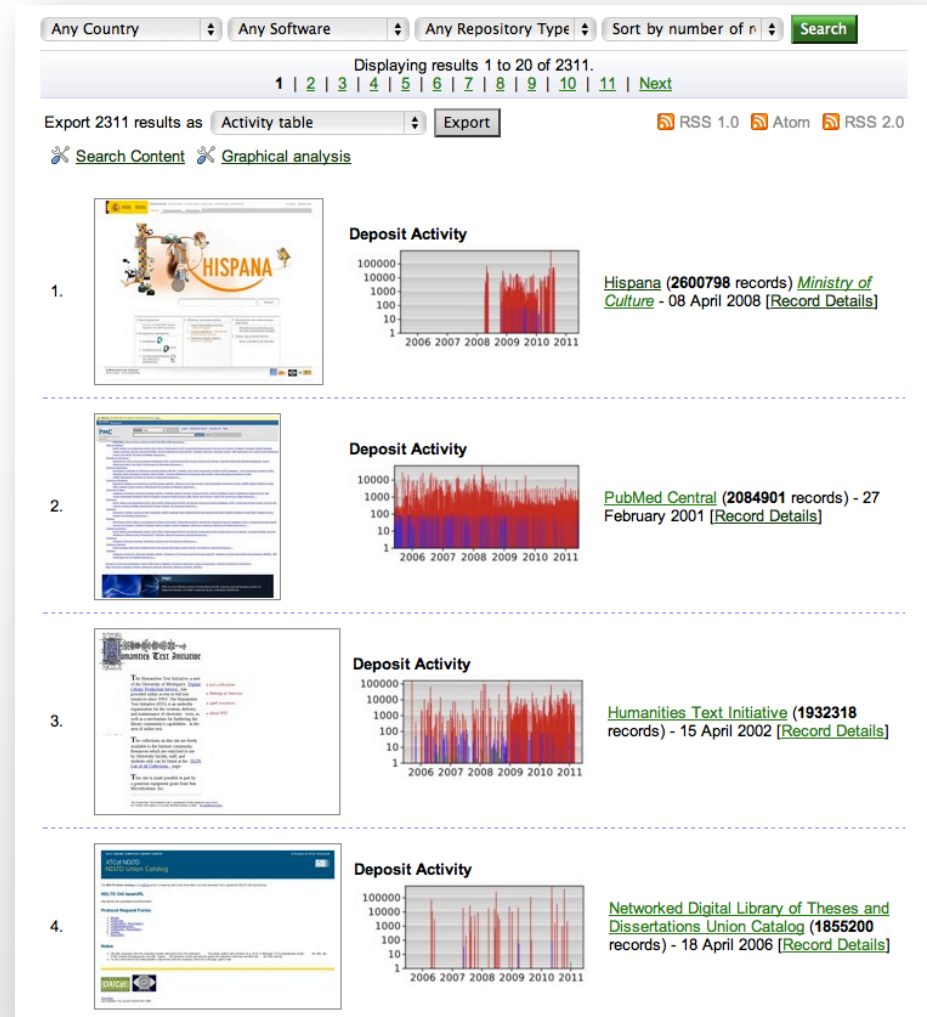
Data -> Big Data

- No controlling data owner
- No defined problem
- No context to interpret

- Raw data? Facts? Objective data?
 - Government employment statistics
 - Higher Education stats e.g. staff/student ratios
 - Sales figures for mobile phones

ROAR: Open Access Big Data

- Evaluating the impact of the web (and repositories) on our global network of information
- Understanding patterns of scholarly information sharing
- How much open access research is being provided?
- Assumes that a repository is just for Open Access research
 - Actually used for many things
 - Count number of deposits
 - May not have any OA files
 - May be pictures or databases
 - May not be research
 - Dates
 - Publication in the repository?
 - Publication in the literature?



Interpreting Data

- E.g. Wikipedia Pages
 - Everyone agrees what a Wikipedia page is for
 - Information is structured / comprehensive
 - DBPedia – enormous internally linked database
 - Users/contributors are hidden
- E.g. Tweets
 - What is Twitter FOR?
 - Updating personal status for friends
 - Sending marketing messages
 - Sharing professional information
 - Information is unstructured / challengingly brief
 - Users are paramount



Andy Smith
@iMinstrel

Facebook: Today I had the happiest day with my wonderful family. I am so blessed.

Twitter: Thank God today is over, I need (lots of) wine!

11/07/2015 08:06

Big Data Computational Thinking

- Twitter is a big database of tweets
 - Users use it to send tweets
 - Users respond to messages
 - Users retweet messages
 - Retweets and responses and information cascades evidence influence
 - We can create metrics and analyse network operation

<http://chrisvoncsefalvay.com/2014/12/07/Gamergate.html>

The Gamergate network isn't a hate network because hate networks don't look like that

Big Data Sociological Thinking

- Twitter is
 - made up of a variety of heterogeneous social practices
 - the digital traces of social life
 - voluntarily proffered opinions
 - at scale, for significant populations
 - evidence
 - attitudes about immigration
 - job losses / searches
 - health practices

Analytic Focus

1. The platform itself

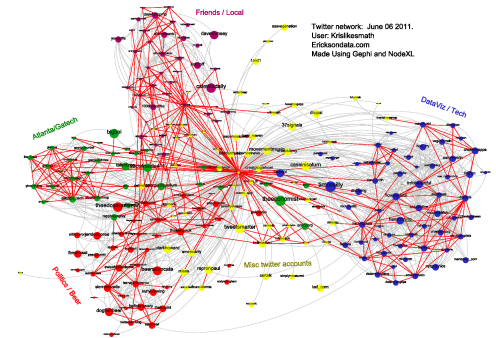
- The abstract networks and linkages
- The data flows and contents

2. The use of the platform

- How did Twitter influence the General Election?
- Does Facebook make you more personally isolated?

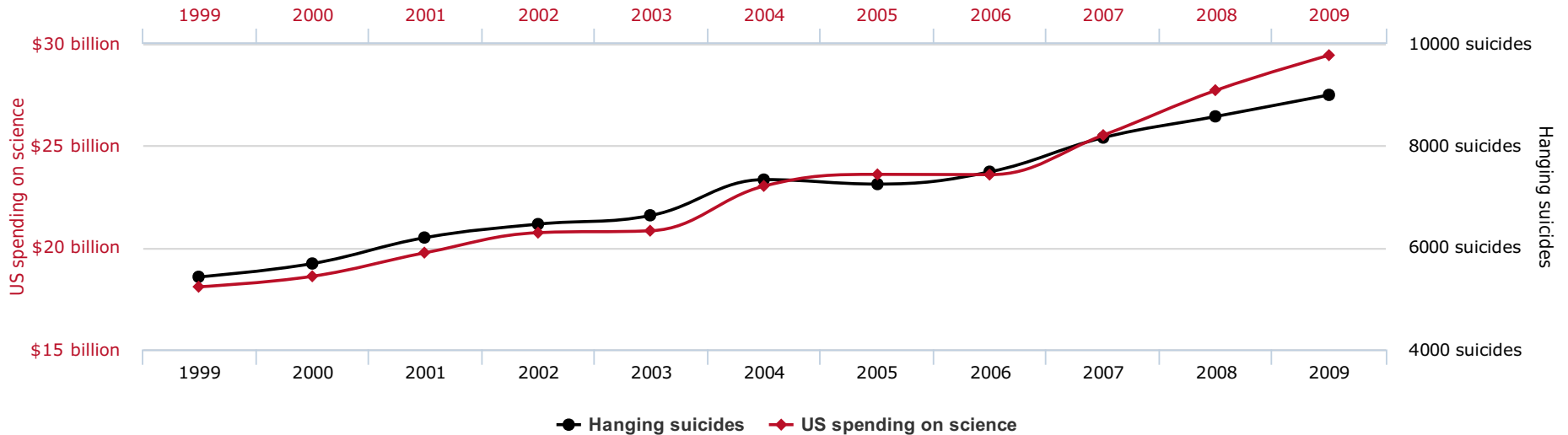
3. The wider world

- Is the UK getting more right wing?
- Do people prefer Coke to Pepsi?



Analytic Methods

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



tylervigen.com

<http://tylervigen.com/>

Spurious Correlations?

- Graphs showing a match over a 10 year period
 - Both data sets progress in step
 - If one increases, so does the other
 - If one decreases so does the other
 - 50% chance each year
 - First year irrelevant
 - Overall chance of match = $2^9 = 1/500$
- Can't just look for patterns!

Correlate?	+	-
+	✓	✗
-	✗	✓



WEB SCIENCE
Institute

UNIVERSITY OF
Southampton

Part II: Web Science Challenges for Social Media Analytics

Susan Halford

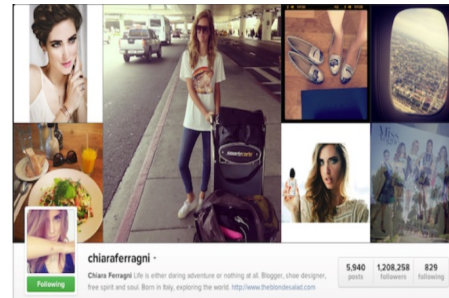
Social Media Analytics

- Big data is not one thing but many ... administrative data, transactional data ...
- Differentiated (amongst other things) by:
 - Content
 - Ownership
 - Structure
 - Availability
- Social media data



Social Media Analytics

- Promise to tell us something about the social:
 - Interactions between individuals, families, friends
 - Social identities
 - Formation of and distinctions between groups
 - Shared meanings and practices
 - Divisions and inequalities



Social Media Analytics

- Promise to tell us something about the social:
 - Scale
 - Proportionality
 - In real time, over time
 - ‘in the wild’
 - Data mining and linking
- A step change in our understanding of the social world?
- How to make the most of this?
- Methodology: *the overall design of research from the conceptualisation of research questions, to methods & tools, tools to data analysis and interpretation*


Methodologies for Social Media Research

- Have social sciences ‘found their telescope’?

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES 

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson  06.23.08



This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behaviour from linguistics to sociology. Forget taxonomy, ontology and psychology. Who knows why people do what they do? ... With enough data the numbers speak for themselves’ Anderson 2008

- Do the numbers speak for themselves?

Methodologies for Social Research

- Conventional practice

Research questions → (Theory) → Methodology → Methods → Data → Analysis

- Social Media Analytics ... sometimes ...

Data → Research Questions → Analysis


- Datasift projects

Research Questions → (Theory) → Data → Methods → Analysis


Methodologies for Social Media Research

- What are the data? #1
 - Ownership and control
 - Platform architectures



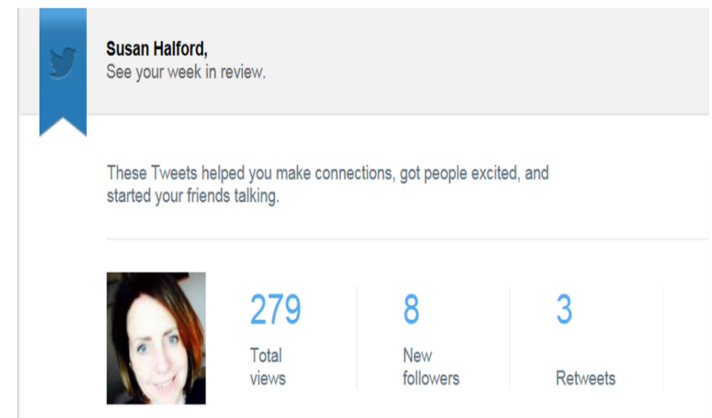
 **Susan Halford,**
See your week in review.

These Tweets helped you make connections, got people excited, and started your friends talking.

	279 Total views	8 New followers	3 Retweets
---	---------------------------	---------------------------	----------------------


Methodologies for Social Media Research

- Technologies shape the data
- Artefacts have histories
- People have practices



Susan Halford,
See your week in review.

These Tweets helped you make connections, got people excited, and started your friends talking.

	279 Total views	8 New followers	3 Retweets
---	---------------------------	---------------------------	----------------------

Methodologies for Social Media Research

- What are the data? #2
 - Demographics

[Tweet](#)

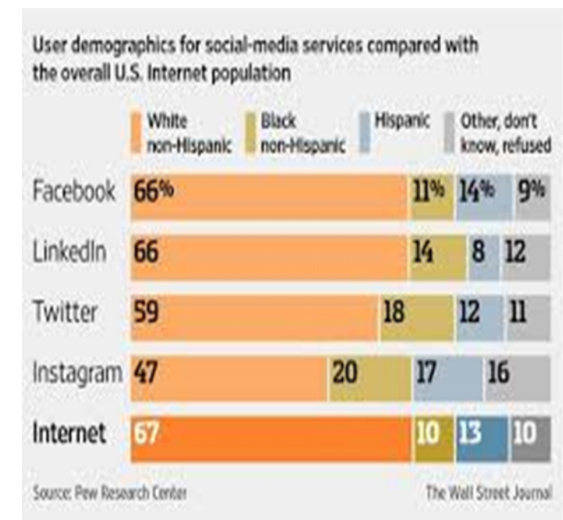
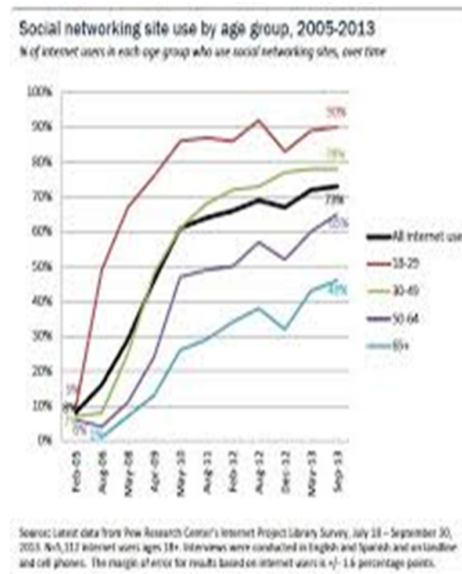
Welcome to Twitter Data

Thursday, July 9, 2015 | By Randy Almond (@rcalmond), Head of Data Marketing [20:44 UTC]

What's possible when you know what the whole world is thinking about any topic at any time? Today we're launching this new blog to showcase the power of Twitter data to help businesses, nonprofits, researchers and the public sector understand the world around them so they can have a greater impact.

Over time, here are a few of the main topics we'll explore:

- **Research.** From understanding public health trends to improving disaster response to uncovering societal insights, we'll highlight the remarkable ways that academics and researchers are analyzing Twitter data. Hear from researchers in the trenches about what they're learning from our data and how they're formulating their findings into meaningful discovery.
- **Visualizations.** Follow [@TwitterData](#) to see some of our data visualizations, and come here to learn the details behind these visualizations, how we created them and what they mean.
- **Data science guidance and insights.** You'll hear from our data scientists about their projects, how they develop ideas for analysis and best practices so you can work with Twitter data.
- **Enterprise-level.** Twitter data is becoming a key element in critical business decisions across enterprise organizations. We'll highlight some of the innovative ways companies are applying



Tweet

Welcome to Twitter Data

Thursday, July 9, 2015 | By Randy Almond (@rcalmond), Head of Data Marketing [20:44 UTC]

What's possible when you know what the whole world is thinking about any topic at any time?

Today we're launching this new blog to showcase the power of Twitter data to help businesses, nonprofits, researchers and the public sector understand the world around them so they can have a greater impact.

Over time, here are a few of the main topics we'll explore:

- **Research.** From understanding public health trends to improving disaster response to uncovering societal insights, we'll highlight the remarkable ways that academics and researchers are analyzing Twitter data. Hear from researchers in the trenches about what they're learning from our data and how they're formulating their findings into meaningful discovery.
- **Visualizations.** Follow [@TwitterData](#) to see some of our data visualizations, and come here to learn the details behind these visualizations, how we created them and what they mean.
- **Data science guidance and insights.** You'll hear from our data scientists about their projects, how they develop ideas for analysis and best practices so you can work with Twitter data.
- **Enterprise-level.** Twitter data is becoming a key element in critical business decisions across enterprise organizations. We'll highlight some of the innovative ways companies are applying

<https://blog.twitter.com/2015/welcome-to-twitter-data>

Methodologies for Social Media Research

- What are the data?
 - Demographics
 - Location?
 - Parody accounts, bots and paid for followers

[Tweet](#)

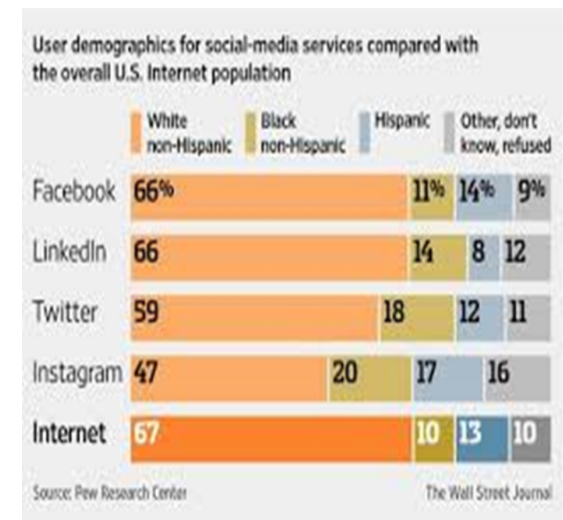
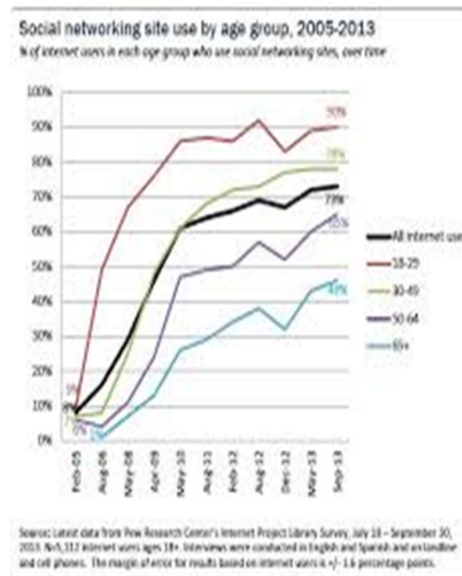
Welcome to Twitter Data

Thursday, July 9, 2015 | By Randy Almond (@rcalmond), Head of Data Marketing [20:44 UTC]

What's possible when you know what the whole world is thinking about any topic at any time? Today we're launching this new blog to showcase the power of Twitter data to help businesses, nonprofits, researchers and the public sector understand the world around them so they can have a greater impact.

Over time, here are a few of the main topics we'll explore:

- **Research.** From understanding public health trends to improving disaster response to uncovering societal insights, we'll highlight the remarkable ways that academics and researchers are analyzing Twitter data. Hear from researchers in the trenches about what they're learning from our data and how they're formulating their findings into meaningful discovery.
- **Visualizations.** Follow @TwitterData to see some of our data visualizations, and come here to learn the details behind these visualizations, how we created them and what they mean.
- **Data science guidance and insights.** You'll hear from our data scientists about their projects, how they develop ideas for analysis and best practices so you can work with Twitter data.
- **Enterprise-level.** Twitter data is becoming a key element in critical business decisions across enterprise organizations. We'll highlight some of the innovative ways companies are applying





TWEETS
5,873

FOLLOWERS
2.17M



Follow

The Dark Lord

@Lord_Voldemort7

Running around leaving scars, collecting my jar of hearts and tearing love apart...

Tweets Tweets & replies Photos & videos



The Dark Lord @Lord_Voldemort7 · 5h

Draco Malfoy is looking rough. pbs.twimg.com/media/B4ScWfEC...



178



221



Thoughts on the royals, politics and much more from Her Majesty, the Twitter

Still Reigning

TWEETS 8,242 FOLLOWERS 1.21M FAVORITES 1

Following

The image shows a Twitter profile for 'Still Reigning'. The banner features a cartoon illustration of Queen Elizabeth II in a purple dress and crown, standing in front of a building. To the left is a yellow circular graphic with the text 'Thoughts on the royals, politics and much more from Her Majesty, the Twitter'. Below the banner is a profile picture of Queen Elizabeth II in a white outfit and hat. The profile statistics are: 8,242 tweets, 1.21M followers, and 1 favorite. A 'Following' button is visible on the right.

Obama Has Millions of Fake Twitter Followers [UPDATED]

5.3k SHARES [Share on Facebook](#) [Share on Twitter](#) +

Ads by Google
[Free eBook for IT Pros](#) - Study Reveals 72% of ITSPs Missing Utilization Targets. Get eBook!
autotax.com



BY ALEX FITZPATRICK

AUG 24, 2012

President Obama has nearly 19 million Twitter followers, but 70%, or approximately 13 million of them, are fake or "inactive," according to a new analysis.

About 30% of Obama's followers — or 5.7 million — were labeled as fake by the analysis, while nearly 40% were found to be inactive.

Mitt Romney's Twitter account, meanwhile, has less than 900,000 followers according to the analysis, but only 15% of them, or about 135,000, are considered fakes and 31%, or about 270,000, are seen as inactive.

The numbers come from Fake Follower Check, a tool from social media firm StatusPeople that analyzes a sampling of a Twitter account's followers and checks for telltale signs of fake followers. Fake accounts are those thought to be created for the sole purpose of sending

Contact Us

SocialMadness

Home Services FAQ Contact About Us

Buy Twitter Followers

Home > Buy Twitter Followers

Worldwide Twitter Followers

Twitter Retweets



Worldwide Twitter Followers

These are Worldwide Twitter Followers. All Followers are of high quality and will be delivered within 2 days maximum.

Choose Amount: 10,000 [+ £45.00]

Enter Full Twitter URL:

Add to Cart

Price: £45.00

Delivery Time: 24-48 Hours

Shopping Cart

Your cart is empty

Discounts

Type SEPT20 at checkout for 20% off your order. No minimum order.

Methodologies for Social Media Research

- Concepts
 - Friendship
 - Identity
- Practices
 - Sarcasm & humour
 - Same tweet, different meanings
 - Platform socialities
 - Embedded meanings
 - Contextual meanings #worstdayever



Gabriella @gaby_colette · 7h

Literally have no wifi for the whole day!! 😞😞

#worstdayever



Bekah Bartlett @bekahbartlett · Jul 2

It's a sad day when @Starbucks discontinues your favorite drink.

#worstdayever #firstworldproblems



Hi I'm Karma @4theluvof7_1_14 · Jul 1

Been a year since I found out I lost my baby #7/1/14 #WorstDayEver



Tim Riggins @VinceSorce21 · Jun 26

Cut my finger at work dropped a chair on my foot and now I just read that gay marriage is legal in all 50 states ... #worstdayever



Weird Mom Ramblings @mom_weird · Jun 23

Our house is pretty much destroyed but we're all safe in a hotel now.

#worstdayever #tornado



View photo

Methodologies for Social Media Research

- Context
 - Ongoing flow of action
 - Temporality
- Demands
 - Cross platform analysis
 - Online/offline data
 - Wide data

Methodologies for Social Media Research

Don't throw the baby out with the bathwater!

Race & Nation

the digital social media networks of pro-and anti-immigration discourse following the lifting of restrictions on Bulgarian and Romanian migration to the UK

Population Estimates

We know already that SNS display distinctive demographic patterns – the question is if and how we can apply appropriate weights to adjust for this.

Platforms for Sikh EDL

the emergent public accommodation of religious minorities inside larger and well established political movements – in this case the Sikh Division of the English Defence League.

OBSEITY & ANOREXIA

recent suggestions regarding the spread of obesity through social networks and considers if and how health interventions might be targeted online

DE Student Summer School Projects