

## Twitter data Collected and Analyzed by Crimson Hexagon

(<http://www.crimsonhexagon.com/>)

**Dates of Data collection** Began 12.00 A.M. GMT 30.03.15 Ended 11.59 P.M. GMT 07.05.15

**Method** – Daily data collection using the Twitter ‘Firehose’ i.e. the universe of Tweets published during this time period. Tweets were captured using the patented algorithmic technology of Crimson Hexagon's platform.

**Sample** - As a first step CH filtered the tweets to select those published in English. Then a second filter was applied to select tweets originating from within in the UK. This was done using geographic information attached to the tweets. Two main methods were employed:

**Geotagged Posts** Some tweets are tagged with the exact coordinates when they are posted. This is possible only if users have tweeted from a mobile phone and have chosen to share their locations. As a result, about 1% of all posts can be located with this kind of precision. These are geotagged posts and can be found by clicking on “Geotagged” and “Geoclusters” on the Geography Beta visualization. It is possible to zoom all the way into the street level. CH selected all tweets that were geotagged as originating within the UK.

**Non-Geotagged Posts** For the remaining 99% of tweets that don't have the exact latitude-longitude coordinates, CH can estimate their locations based on various pieces of information such as their profile information. Based on the very large corpus of tweets CH has available it is possible to identify the attributes of users whose locations are known, and match users whose locations are not directly available but display similar attributes. Based on the 1% of all tweets that are geotagged, which is hundreds of millions of tweets from across the globe CH have developed statistical guesser for estimating location. The most useful piece of information in this process is the “location” field in user profiles, which is a free-form input space where users can describe their locations in their own words. In addition, other attributes such as users' time zones and languages can be used to determine the countries, regions and cities from where they have tweeted. If a match is found, our guesser estimates the location of the post based on this similarity. When no match is found, we simply do not make a prediction. The locations of such posts are labelled as unknown.

The more precise the location to be estimated, then inevitably the harder it is to find a good match. In general, we are able to match:

- About 90% of all posts to a country of origin
- About 70% of all posts to a specific state or province within that country
- About 50% of all posts to a city or urban area within that state

N.B. The CH algorithm specifically excludes Tweets if there is insufficient information to determine with confidence that a post was from a specific location.

In the case of the UK election data collection project 359,395,390 English language posts were identified from the overall UK conversation on Twitter (30 Mar - 7 May) which represented an average of 9,215,266 English Tweets per day from the UK (30 Mar - 7 May). Within this overall total,

32,578,454 Tweets fell within our analysis parameters i.e. mentioned one of our key terms of interest and our reports were compiled on a daily average of 835,345 Tweets falling within our analysis parameters.

**Data Dictionary Search terms** – After filtering to create our sub-sample of tweets we then applied a series of keywords as search terms to create our issue categories and election specific terms. These were:

## **ISSUE CATEGORIES**

Example:-

Banks – Bank of England: "Bank of England" OR bankofengland

Banks – Bank Regulation: "bank regulation" OR "bank regulations" OR "bank regulating" OR "bank regulators" OR "Bank Regulator" OR bankregulation OR bankregulations OR bankregulating OR bankregulators OR bankregulator

## **ELECTION SPECIFIC TERMS**

Example:-

Conservatives – Conservatives Party:"Conservative Party" OR conservativeparty

Attached is an appendix showing the Data Dictionary issue categories and the search terms. Below that is the full list of key words and their variations.

For leader all mentions just the mentions of the biggest six parties are included due to certain names being omitted from the search due to the higher likelihood of false positive (for example, Leanne Wood and Peter Robinson). For leader @mentions, all parties' leaders are included. Similarly, for party mentions and party @mentions, all parties are included.

The overall N of unique mentions in the dataset is 41, 827,928. This is lower than the total mentions of 44, 441,797. The discrepancy arises from the fact that we report separately all mentions of proper names for parties and leaders and all specific @mentions of their twitter handle. Given that all @mentions are automatically included in the 'all leader mentions' and 'all party mentions' columns, we remove the @mentions when calculating the total unique mentions to avoid double counting. There are four cases where this 'rule' is not followed for the leaders. In the case of Leanne Wood, Peter Robinson, Gerry Adams and George Galloway we did not include their proper names as search terms given they were considered more common names which could increase the chance of false positives and inflate the overall count. They were, however, left in for @mentions as the process for @mentions searches the exact handle and so avoids this potential problem. Thus in calculating unique mentions, the @mentions were included for these four leaders.

<b>Banks</b>	<b>Consumer Debt</b>	<b>Crime</b>	<b>Economy</b>	<b>Education</b>	<b>Energy</b>
Bank of England	Credit Card Debt	Break the Law	#austeritybritain	Curriculum	Electricity
Bank Regulation		Corruption	#autumnstatement	Ofsted	Energy Prices
Bonuses		Courts	#costoflivingcrisis	Tuition	Gas Prices
Finance Regulation		Criminal Offence	#thedeficit	University Fees	
Finance Service		Law and Order	#wastenotwantnot		
Lending		Law Breaker	Austerity		
Mark Carney		Law Enforcement	Autumn Statement		
Tax Dodging		National Crime Agency	Budget Deficit		
		NCA	Budget Surplus		
		Police	Chancellor Statement		
			Cost of Living		
			Deficit		
			Employment		
			GDP		
			Interest Rates		
			National Debt		
			Privatisation		
			Public Debt		
			Recession		
			Small Business		
			Treasury		





<b>Unemployment</b>	<b>Welfare</b>	<b>Election@</b>	<b>Election@</b>	<b>Election#</b>	<b>Election#</b>
#jobs	Benefit Cheats	@allianceparty	@RespectParty	#allianceparty	#SDLP
#unemployment	Benefit Cuts	@bnp	@RespectPartyUK	#BattleForNumber10	#SDLP
Active Population	Benefit Fraud	@BritishNational	@sdlp	#brexit	#SinnFein
Apprenticeship	Welfare Cheats	@Conservatives	@sdlplive	#conservatives	#SinnFein
Dole	Welfare Fraud	@David_Cameron	@sinnfein	#Devolution	#SNP
Fewer Jobs		@dup	@sinnfeinireland	#DUP	#SNP
Job Creation		@duponline	@snp	#GE15	#ukelection
Job Cuts		@Ed_Miliband	@TheGreenParty	#GE2015	#ukelection
Job Seekers		@georgegalloway	@theSNP	#generalelection	#UKIP
Jobs Cut		@GerryAdamsSF	@toryes	#GreenParty	#UKIP
Lack of Jobs		@uklabour	@UKIP	#greensurge	#votegreen2015
More Jobs		@leannewood	@UKLabour	#labour	#VoteGreen2015
Unemployment		@libdem	@uup	#labourdoorstep	
Wage Gap		@LibDems	@uuponline	#LD2015	
Zero Hours		@nick_clegg		#leadersdebate	
		@NicolaSturgeon		#plaid15	
		@Nigel_Farage		#plaidcymru	
		@peterrobinson		#plaidnation	
		@plaid		#plaidsurge	
		@Plaid_Cymru		#plaidsurge	

<b>Election Terms</b>	<b>Election Terms</b>
Alex Salmond	Natalie Bennett
Alliance Party	natalieben
BNP	Nick Clegg
British National Party	Nicola Sturgeon
Cameronettes	Nigel Farage
Coalition	Plaid Cymru
Caroline Lucas	Polling Station
Conservative Party	Polling Day
Conservatives	Respect Party
David Cameron	Scottish National Party
Democratic Unionist Party	SDLP
DUP	Sinn Fein
Ed Miliband	SNP
Election Debate	Social Democratic and Labour Party
General Election	Tories
Gerry Adams	Tory Party
Green Party	UKIP
Labour	UK Independence Party
Labour Party	UUP
Leader Debate	
Lib Dems	
Liberal Democrats	
Milifandom	

Milifans	

Name	Search Term	keywords
Banks	Bank of England Bank Regulation Bank Regulation Bank Regulation Bonuses Finance Regulation Finance Regulation Finance Regulation Finance Regulation Finance Service Finance Service Lending Mark Carney Tax Dodging Tax Dodging	"Bank of England" OR bankofengland "bank regulation" OR "bank regulations" OR "bank regulating" OR "bank regulators" OR "Bank Regulator" bankregulating OR bankregulators OR bankregulatoOR bankregulation OR bankregulations OR bankregulating OR bankregulators OR bankregulator Bonuses "Finance Regulation" OR "Finance regulations" OR "finance regulating" OR "Finance regulators" OR "finance regulator" OR "financial regulations" OR "financial regulating" OR "financial regulators" OR "financial regulator" OR financeregulation OR "financial regulation" OR financialregulator OR financeregulator OR financialregulation OR financialregulations OR financialregulating OR financialregulators "finance service" OR "finance services" OR "financial service" OR "financial services" OR financeservice financialservices OR financeservices OR financialservice Lending "Mark Carney" OR markcarney OR "Mark Carneys" OR markcarneys "tax dodge" OR "tax dodges" OR "tax dodging" OR "tax dodger" OR "tax dodgers" OR taxdodge OR taxdodges OR taxdodging OR taxdodger OR taxdodgers
Consumer Debt	Credit Card Debt	"credit card debt" OR "credit card debts" OR "creditcard debt" OR "creditcard debts" OR creditcarddebt OR creditcarddebts
Crime	Break the Law Corruption Courts Criminal Offence Law and Order	"Break the law" OR "Broke the law" OR "Breaking the law" OR breakthelaw OR brokethelaw OR breakingthelaw Corruption Courts OR Court "Criminal Offence" OR "Criminal Offences" OR criminaloffence OR criminaloffences "Law and Order" OR lawandorder OR "Law & Order" OR law&order



Law Breaker	"Law Breaker" OR "Law breakers" OR lawbreaker OR lawbreakers
Law Enforcement	"Law Enforcement" OR "Law Enforcements" OR lawenforcement OR lawenforcements
National Crime Agency	"National Crime Agency" OR "National Crime Agency's" OR "National Crime Agencies" OR nationalcrimeagency
National Crime Agency	OR nationalcrimeagency's OR nationalcrimeagencies
NCA	NCA OR NCAs
Police	Police

Economy	#austeritybritain	#austeritybritain
	#autumnstatement	#autumnstatement
	#costoflivingcrisis	#costoflivingcrisis
	#thedeficit	#thedeficit
	#wastenotwantnot	#wastenotwantnot
	Austerity	Austerity
	Autumn Statement	"Autumn Statement" OR "Autumns Statement" OR "Autumn Statements" OR "Autumns Statements" OR autumnstatement
	Autumn Statement	OR autumnsstatement OR autumnstatements OR autumnsstatements
	Budget Deficit	"Budget Deficit" OR "Budget Deficits" OR budgetdeficit OR budgetdeficits
	Budget Surplus	"Budget Surplus" OR budgetsurplus
	Chancellor	
	Statement	"Chancellor Statement" OR "chancellors statement" OR "chancellor statements"
	Chancellor	
	Statement	OR chancellorstatements OR chancellorsstatements
	Chancellor	
	Statement	OR "chancellors statements" OR chancellorstatement
	Cost of Living	"Cost of Living" OR "Costs of living" OR costofliving OR costsofliving
	Deficit	Deficit OR deficits
	Employment	Employment OR employments
	GDP	GDP OR GDPs
	Interest Rates	"Interest Rates" OR "Interest Rate" OR interestrates OR interestrates

	National Debt Privatisation Public Debt Recession Small Business Treasury	"National Debt" OR "National Debts" OR nationaldebt OR nationaldebts Privatisastion OR privatisations "public debt" OR "public debts" OR publicdebt OR publicdebts recession OR recessions "Small Business" OR "small businesses" OR smallbusiness OR smallbusinesses treasury
--	--	---

Education	Curriculum Ofsted Tuition University Fees	curriculum OR curriculums Ofsted Tuition OR Tuitions "University Fees" OR "University Fee" OR "Universities Fees" OR universityfees OR universityfee OR universitiesfees
-----------	--	---



Energy	Electricity Energy Prices Gas Prices	Electricity "Energy Prices" OR "Energy price" OR energyprices OR energyprice "Gas Prices" OR "Gas Price" OR gasprices OR gasprice
Environment	Climate Change Flooding Fracking Global Warming Overpopulation Pollution Renewable Energy Wind Farms	"Climate Change" OR climatechange Flooding Fracking "Global Warming" OR globalwarming Overpopulation OR "Over population" Pollution "Renewable Energy" OR "Renewable Energies" OR renewableenergy OR renewableenergies "Wind Farms" OR "wind farm" OR windfarms OR windfarm
Europe	@EU_Commission @eucopresident @JunckerEU #TeamJunckerEU Barroso EU European Arrest Warrant European Commission European Council European Laws European Legislation European Parliament Herman Van Rompuy Jean Claude Juncker MEPs	@EU_Commission @eucopresident @JunckerEU #teamjunckerEU Barroso EU "European Arrest Warrant" OR "European Arrest Warrants" OR europeanarrestwarrant OR europeanarrestwarrants "European Commission" OR "European Commissions" OR europeancommission OR europeancommissions "european council" OR "european councils" OR europeancouncil OR europeancouncils "European Laws" OR "european law" OR europeanlaws OR europeanlaw "European Legislation" OR "European Legislations" OR europeanlegislation OR europeanlegislations "European Parliament" OR europeanparliament "Herman Van Rompuy" OR "Herman Van Rompuys" OR hermanvanrompuy OR hermanvanrompuys "Jean Claude Juncker" OR "Jean Claude Junckers" OR jeanclaudejuncker OR jeanclaudejunckers MEPs
Fuel Prices	Petrol	Petrol OR petrols

Housing	Affordable Housing House Prices Housing Bubble Housing Crisis Housing Tax Mortgage New Builds New Homes Property Bubble Property Price Property Tax Stamp Duty	"Affordable Housing" OR affordablehousing "House Prices" OR houseprices "Housing Bubble" OR housingbubble "Housing Crisis" OR housingcrisis "Housing Tax" OR "Housing Taxes" OR housingtax OR housingtaxes Mortgage OR mortgages "New Builds" OR newbuilds "New Homes" OR newhomes "property bubble" OR propertybubble "property price" OR "property prices" OR propertyprice OR propertyprices "Property Tax" OR propertytax OR "Property taxes" OR propertytaxes "Stamp Duty" OR stampduty
Immigration	Asylum Border Control Foreigners Home Office Illegal Immigration Immigrant Racism Refugee	Asylum "Border Control" OR bordercontrol Foreigners OR Foreigner "home office" OR "home offices" OR homeoffice OR homeoffices "Illegal Immigration" OR "Illegal immigrant" OR "illegal immigrants" OR illegalimmigration OR illegalimmigrant OR illegalimmigrants immigrant OR immigrants OR immigration Racism refugee OR refugees
Independence	#evel BreakupoftheUK Constitutional Reform Devo Max Devolution English Parliament English Votes for English Laws EVEL Referendum	#evel BreakupoftheUK OR "Break up of the UK" "Constitutional Reform" OR "Constitutional Reforms" OR constitutionalreform OR constitutionalreforms "Devo Max" OR devomax Devolution "English Parliament" OR englishparliament "English votes for english laws" OR englishvotesforenglishlaws EVEL Referendum OR referendums

Inflation	Cost of Living Rising Prices	"Cost of Living" OR "Costs of living" OR costofliving OR costsofliving "Rising Prices" OR "rising price" OR risingprices OR risingprice
International Problems	Afghanistan Ebola Iraq ISIL ISIS Islamic Fundamentalism Islamic State Russia Syria Tikrit Massacre Ukraine	Afghanistan OR Afghanistans Ebola OR Ebolas Iraq OR Iraqs ISIL ISIS "Islamic fundamentalism" OR islamicfundamentalism "Islamic State" OR islamicstate Russia OR Russias Syria OR Syrias "Tikrit Massacre" OR tikritmassacre Ukraine OR Ukraines
NHS	#111linecrisis #Health #NHS #NHS111 #NHScrisis #saveournhs A&E Healthcare TTIP	#111linecrisis #Health #NHS #NHS111 #nhscrisis #saveournhs "A&E" OR "A and E" OR aande Healthcare OR "health care" TTIP
Pensions	State Pension	"State Pension" OR "State Pensions" OR statepension OR statepensions

Poverty	Child Poverty Living Standards Living Standards Living Wage Low Pay Low Wages Minimum Wage Rising Prices Social Inequality	"Child Poverty" OR childpoverty "living standards" OR "living standard" OR "standard of living" OR "standards of living" OR livingstandards OR livingstandard OR standardofliving OR standardsofliving "Living Wage" OR livingwage "Low Pay" OR Lowpay "Low Wages" OR "Low Wage" OR lowwages OR lowwage "Minimum Wage" OR minimumwage "Rising Prices" OR "rising price" OR risingprices OR risingprice "Social Inequality" OR socialinequality
Taxes	Tax	Tax OR Taxes
Terrorism	GCHQ MI5 MI6 Terror Threat	GCHQ MI5 MI6 "Terror Threat" OR "Terror threats" OR terrorthreat OR terrorthreats
Unemployment	#jobs #unemployment Active Population Apprenticeship Dole Fewer Jobs Job Creation Job Cuts Job Seekers Jobs Cut Lack of Jobs More Jobs Unemployment Wage Gap Zero Hours	#jobs #unemployment "Active Population" OR activepopulation Apprenticeship OR appretinceships Dole "fewer Jobs" OR fewerjobs "Job Creation" OR jobcreation "Job Cuts" OR jobcuts "Job Seekers" OR "Job seeker" OR jobseekers OR jobseeker "Jobs cut" OR "Job cuts" OR jobscut OR jobcuts "Lack of Jobs" OR lackofjobs "more jobs" OR morejobs Unemployment "Wage Gap" OR "Wage gaps" OR wagegap OR wagegaps "Zero Hours" OR "Zero Hour" OR zerohours OR zerohour

Welfare	Benefit Cheats Benefit Cuts Benefit Fraud Welfare Cheats Welfare Fraud	"Benefit Cheats" OR "Benefit Cheat" OR benefitcheats OR benefitcheat "Benefit Cuts" OR benefitcuts "Benefit Fraud" OR benefitfraud "Welfare Cheats" OR "Welfare Cheat" OR welfarecheats OR welfarecheat "Welfare Fraud" OR "Welfare Frauds" OR welfarefraud OR welfarefrauds
---------	--	--



Election DD	@allianceparty @bnp @BritishNational @Conservatives @David_Cameron @dup @duponline @Ed_Miliband @georgegalloway @GerryAdamsSF @labour @leannewood @libdem @LibDems @nick_clegg @NicolaSturgeon @Nigel_Farage @peterrobinson @plaid @Plaid_Cymru @RespectParty @RespectPartyUK @sdlp @sdlplive @sinnfein @sinnfeinireland @snp @TheGreenParty @theSNP	@allianceparty @bnp @BritishNational @Conservatives @David_Cameron @dup @duponline @Ed_Miliband @georgegalloway @GerryAdamsSF @labour @leannewood @libdem @LibDems @nick_clegg @NicolaSturgeon @Nigel_Farage @peterrobinson @plaid @Plaid_Cymru @RespectParty @RespectPartyUK @sdlp @sdlplive @sinnfein @sinnfeinireland @snp @TheGreenParty @theSNP
-------------	--	--

@tories @UKIP @UKLabour @uup @uuponline #allianceparty #BattleForNumber10 #brexit #conservatives #Devolution #DUP #GE15 #GE2015 #generalelection #GreenParty #greensurge #labour #labourdoorstep #LD2015 #leadersdebate #plaid15 #plaidcymru #plaidnation #plaidsurge #plaidsurge #SDLP #SDLP	@tories @UKIP @UKLabour @uup @uuponline #allianceparty #BattleForNumber10 #brexit #conservatives #Devolution #DUP #GE15 #GE2015 #generalelection #GreenParty #greensurge #labour #labourdoorstep #LD2015 #leadersdebate #plaid15 #plaidcymru #plaidnation #plaidsurge #plaidsurge #SDLP #SDLP
---	---

#SinnFein	#SinnFein
#SinnFein	#SinnFein
#SNP	#SNP
#SNP	#SNP
#ukelection	#Ukelection
#ukelection	#Ukelection
#UKIP	#UKIP
#UKIP	#UKIP
#votegreen2015	#votegreen2015
#VoteGreen2015	#VoteGreen2015
#votegreen2015	#votegreen2015
#VoteGreen2015	#VoteGreen2015
Alex Salmond	"Alex Salmond" OR "Alex Salmonds" OR alexsalmond OR alexsalmonds
Alex Salmond	"Alex Salmond" OR "Alex Salmonds" OR alexsalmond OR alexsalmonds
Alliance Party	"Alliance Party" OR allianceparty
Alliance Party	"Alliance Party" OR allianceparty
BNP	BNP
British National Party	"British National Party" OR britishnationalparty
Caroline Lucas	"Caroline Lucas" OR carolinelucas
Caroline Lucas	"Caroline Lucas" OR carolinelucas
Conservative Party	"Conservative Party" OR conservativeparty
Conservatives	Conservatives
David Cameron	"David Cameron" OR davidcameron OR "David Camerons" OR davidcamerons
David Cameron	"David Cameron" OR davidcameron OR "David Camerons" OR davidcamerons
Democratic Unionist Party	"Democratic Unionist Party" OR democraticunionistparty
DUP	DUP

Ed Miliband	"Ed Milliband" OR "Ed Miliband" OR "Ed Millibands" OR "Ed Milibands" OR edmilliband OR edmiliband OR edmillibands OR edmilibands
Ed Milliband	"Ed Milliband" OR "Ed Miliband" OR "Ed Millibands" OR "Ed Milibands" OR edmilliband OR edmiliband OR edmillibands OR edmilibands
Election Debate	"Election debate" OR "Election Debates" OR electiondebate OR electiondebates
Election Debate	"Election debate" OR "Election Debates" OR electiondebate OR electiondebates
General Election	"General Election" OR generalelection
General Election	"General Election" OR generalelection
Gerry Adams	"Gerry Adams" OR gerryadams
Gerry Adams	"Gerry Adams" OR gerryadams
Green Party	"Green Party" OR greenparty
Labour	Labour
Labour Party	"Labour Party" OR labourparty
Leader Debate	"leader debate" OR "leaders debate" OR "Leader debates" OR "Leaders debates" OR leaderdebate OR leadersdebate
Leader Debate	OR leaderdebates OR leadersdebates
Lib Dems	"lib dems" OR libdems
Liberal Democrats	"Liberal Democrats" OR liberaldemocrats
Natalie Bennett	"Natalie Bennett" OR "Natalie Bennetts" OR nataliebennett OR nataliebennetts
Natalie Bennett	"Natalie Bennett" OR "Natalie Bennetts" OR nataliebennett OR nataliebennetts
natalieben	@natalieben
natalieben	@natalieben
Nick Clegg	"Nick Clegg" OR "Nick Cleggs" OR nickclegg OR nickcleggs
Nick Clegg	"Nick Clegg" OR "Nick Cleggs" OR nickclegg OR nickcleggs
Nicola Sturgeon	"Nicola Sturgeon" OR "Nicola Sturgeons" OR nicolasturgeon OR nicolasturgeons
Nicola Sturgeon	"Nicola Sturgeon" OR "Nicola Sturgeons" OR nicolasturgeon OR nicolasturgeons
Nigel Farage	"Nigel Farage" OR "Nigel Farages" OR nigelfarage OR nigelfarages
Nigel Farage	"Nigel Farage" OR "Nigel Farages" OR nigelfarage OR nigelfarages

Plaid Cymru	"Plaid Cymru" OR plaidcymru
Respect Party	"Respect Party" OR respectparty
Respect Party	"Respect Party" OR respectparty
Scottish National Party	"Scottish National Party" OR scottishnationalparty
SDLP	SDLP
Sinn Fein	"Sinn fein" OR sinnfein
SNP	SNP
Social Democratic and Labour Party	"Social Democratic and Labour Party" OR "Social Democratic & Labour Party"
Social Democratic and Labour Party	OR socialdemocraticandlabourparty OR "socialdemocratic&labourparty"
Tories	Tories
Tory Party	"Tory Party" OR toryparty
UK Independence Party	"UK Independence Party" OR Ukindependenceparty
UKIP	UKIP
Ulster Unionist Party	"Ulster Unionist Party" OR ulsterunionistparty
UUP	UUP