French Social Media Data – Exercise

Background

The data in the spreadsheet are a subset of a larger number of fields captured from Twitter, Facebook, Reddit and Digg as a test during the French Presidential Election of 2012 – see

https://en.wikipedia.org/wiki/French_presidential_election,_2012 for the contestants.

What can you learn from the data with some simple manipulation in Excel?

Examine the spreadsheet

- 1. How many fields (columns) are there?
- 2. How many records (rows) are there?
- 3. What do you notice about the data (clues below)?

Do all the fields contain values?

Many of the fields are only partially populated (i.e. containing records) with plenty of blank, or NULL, values. How can you count these to get a feel for sparsity across the dataset?

Tip: Use the Excel function COUNTBLANK to calculate the number of NULL values in each field. Go to the bottom of the first column (**interaction_id**) by hitting CTRL+DOWN ARROW; in the blank cell beneath the last record type the formula =COUNTBLANK(A2:A52915), you should see the value 0. Highlight this cell and place the mouse cursor over the bottom right hand corner until you see a little cross. Click on this cross and pull it rightwards until you've copied the formula across all the columns in the spreadsheet.

Question: Can you calculate the number of blanks as a percentage of the total and produce a meaningful graph?

AutoSave 🖲 Off) 🖪 🕤 T 🔿 - 👎								fred - Ex	cel							Adria	n Tear 🛛 🗈	-	٥	×
File Home Insert Page Layout F	ormulas	Data Rev	iew Vi	ew Help	Ω Tell π	e what you v	want to do													Share
A Cut Calibri • 11 Paste ▲ Copy • Paste ▲ Compt Painter	• A A		≫- € →	Wrap T	ext & Center 👻	General	• 00, 0, 0 0, 00, 0	Conditiona	I Format as	Normal Neutral	Ba	d Iculation	Good Check (Cell	Insert	Delete Format	∑ AutoSum ↓ Fill +	Sort &	Find &	
Cichard Fanter	_				_	b lumb		Formatting	 Table * 		Ch los					· ·	· erear	Filter - S	select -	
Cipboard is Font	19		Aligni	nent	19	Numbi	6L i A		3)	by hitting C	Styles TRI +DOWN	N ARROW: i	n the blank	cell		Cens	t	aiting		^
j n ▼ ×				fred								, ou should	see the va	lue 0.						
A	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R S	T		
1 interaction_id	interactior i	interactior in	nteractio	interactio	interaction	demograp	klout_sco	interaction	interaction	twitter_m t	witter_us t	witter_ret	witter_re tv	vitter_re sa	lience_c	interaction inte	raction_geo_	longitude		
2 0	0	0	40	21	0	20694	1521	0	21	45018	26010	27820	27597	27597	8816	52178	52178			_
3																				
4 interaction_id	0	0.00%									Chart	Tiele								
5 interaction_type	0	0.00%									Chart	litle								
6 interaction_created_at	0	0.00%				interaction m	huttensi													
7 interaction_source	40	0.08%				interaction_g	eo_iongitua	e												
8 interaction_author_username	21	0.04%				interaction	_geo_latitud	e												_
9 interaction_author_name	0	0.00%			58	lience_conte	nt_sentimen	t												
10 demographic_gender	20694	39.11%				twitter_re	etweet_coun	t												
11 klout_score	1521	2.87%				twitter_	retweeted_i	b												
12 interaction_content	0	0.00%			twitter_retw	eet_user_fol	lowers_coun	t												
13 interaction_link	21	0.04%			twi	tter user fol	lowers cour					_								
14 twitter_mentions	45018	85.08%				twitt	ter mention								_					_
15 twitter_user_followers_count	26010	49.16%				CWIC	ter_mention													_
16 twitter_retweet_user_followers_count	27820	52.58%				int	teraction_lin	k												_
17 twitter_retweeted_id	27597	52.15%				interac	tion_conten	t												
18 twitter_retweet_count	27597	52.15%					klout_scor	e 🔳												
19 salience_content_sentiment	8816	16.66%				demogra	aphic_gende	r												_
20 interaction_geo_latitude	52178	98.61%				interaction_	author_nam	e												_
21 interaction_geo_longitude	52178	98.61%			inte	raction auth	or usernam	p												
22						inters	action source													_
23						intere	action_aoure													_
24						interactio	n_created_a	t												_
25						inte	eraction_typ	e												_
26							nteraction_i	d												_
27								0.00%	20.0	0%	40.00%	6	0.00%	80.00%	6	100.00%	120.00	%		_
28																				
29																				
Sheet1 % Unpopulated	Time histo	ogram	(+)								4									Þ
Ready	_											Averag	ge: 16083.9444	14 Count: 36	Sum: 289	511 🏢 🛙	1 🗉 – —	-	+	100%

Tip: You'll need to know the total number of records, create another formula and – ideally – copy and transpose (Paste Special...) these values into a new sheet. Use the plus button inside a circle at the bottom of Excel to create the new sheet; select the first row and the COUNTBLANK row you created above. Hit CTRL+C to copy and paste these values into a new sheet. Select them then hit CTRL+C again then go to the Home ribbon, Paste, Paste Special, Transpose. Use a formula to create percentages and insert a graph. Your solution may look something like the above.

Question: Can you order the graph so the most populated columns are shown first? Does it make more sense to show the %populated or the %unpopulated?

How are the records distributed in time?

- 1. When was the first record created?
- 2. When was the last record created?
- 3. How much time has elapsed?

The field **interaction_created_at** contains record creation date/time (and timezone offset). Unfortunately, Excel cannot understand the human readable format...

Tip: Click on the column heading to select all rows in this column, hit CTRL+C to copy these into a new sheet (click the cross within the circle at the bottom of Excel to create one). In cell B1 type a heading for your new column (e.g. **creationtime**). In B2 devise a formula that will turn the date/time literal into a time data type. Look at the first record:

Sun, 06 May 2012 16:17:47 +0000 123456789012345678901

The numbers below the date/time show character positions 1 through 10 (shown as 0); the field length is 31 overall. You can use the function MID to select parts of this field and pass days, months and years into Excels TIME function using the formula:

=TIME(MID(A2,18,2),MID(A2,21,2),MID(A2,24,2))

Work out what's going on! The TIME function expects hours, minutes and seconds. See how the MID function is used to select these (e.g. starting at character 18 reading 2 characters to get hours) from the long (human readable) date/time with timezone offset format and convert them into something Excel can understand.



Once you've got your formula working select the cell you created it in and double click the little cross on the bottom right hand corner; this will copy the formula all the way down to the bottom of the sheet. Select your **creationtime** column and from the Insert ribbon insert a Histogram. It should look something like the above.

Can any of the Tweets be mapped?

You will notice the spreadsheet contains two fields/columns named **interaction_geo_latitude** and **interaction_geo_longitude** – some users, when they Tweet, agree to post their geographic coordinates alongside their message. These coordinates are saved from the Global Positioning System (GPS) chip in their phones etc.

Excel does not have great capabilities for mapping from Latitude and Longitude, but this can be overcome! Copy the two columns named above into another new sheet. It should look something like this:

A	utoSave 💽 🕅 🚍	ڻ - ر ک							fred - Exce	I							Adrian	Tear 🗈	ı –	٥	×
	le Home Inser	t Pagel	ayout Formulas	Data Rev	riew View	Help 🖓	ell me what yo														Share
Past	Cut	Calibri	• 11 • A A	= = =	≫- ₽ 	Wrap Text	General		Conditional	ormat as	Normal	Bad	lation	Good	- -	Insert D	elete Format	∑ AutoSi ↓ Fill -	um * AZ▼ Sort &	Find &	
*	✓ Format Painter	D I U	· · · · · · · · ·			werge & cente	/0	· .00 →.0	Formatting -	Table -	veutrai	Calco	alation	CHECK C		*		Clear •	Filter -	Select +	
	Clipboard Is	- ×	Font		Alignment	fred	Nur	nber 5		a) by	hitting CTF	Styles	ARROW; in ou should s	the blank o see the valu	ell ue 0.		Cells		Editing		^
	A		в	C	D	F	F G	н			к.		M	 N	0	Р	0	R	S	т	
1	interaction_geo_lati	tude inte	raction_geo_longitud	le	-	-											~				70
2																					
3																					
4																					
5																					
7																					
8																					
9																					
10																					
11																					
12																					
13																					
14																					
15	42 5440	19105	1 2/20	12																	
17	45.5440	0155	1.5425	12																	
18																					
19																					
20																					
21																					
22																					
23																					
24																					
26																					
27																					
28																					
29																					-
	Sheet1	% Unpop	oulated Time histo	gram C	reate map	\oplus						4									Þ
Read	у											Av	rerage: 89.272	00842 Cour	t: 1476 Sum	: 131586.940	4 🎹 🗉) E -	- F	+	100%
																					_

Then copy/paste the interaction_content column into this sheet as well, it should look like this:

AutoSav	e 💽 Off	5 •∂•∓												
File	Home Ins	ert Page Layout	Formulas Dat	ta Review Vie	w Help ⊽⊺e	ll me what you want to do							Ŀ	ත් Share
Paste	Cut Copy ÷ Format Painter	Calibri • 11 B I U •	• <u>∆</u> • <u>A</u> • ≡	= = ⊗ . = = € €	Wrap Text	General	Conditional Format as Formatting - Table -	Normal Neutral	Bad Calculation	Good Check Cell	Insert Delete Form	→ ∑ AutoSum → Fill → mat → Clear →	Sort & Find Filter - Select) & :t -
Clipb	oard G	▼ ×	G.	Alignm	fred	Number 🕞	le Ie	by hitting CT	Styles RI +DOWN ARROW; , ou shoul	in the blank cell d see the value 0.	Cells		Editing	^
	A		В								C			^
1 interac	ction_geo_la	titude interaction_	geo_longitude	interaction_conte	ent									
2				nttp://t.co/Sivici	PXW Sarkozy, Hollai	nde : qui sera le prochail	h prA©sident ?: La part	icipation prog	gresse par rapport	nttp://t.co/n8Atw/mw				
2				Els primers sondat	ges atorguen la vict	A-ria a Hollande - Vilavi	eb http://t.co/bxen1/t	w va-a @viia	aweb					
4				PT @purtapa: Atto	ent nate de voir noi andos lo rã@sultat s	lande a l'oeuvre x) - Ame	en- et surtout allez voter iur	aulã la domi	iñ"ro cocondo "#boll:	ando				
6				La France devient	Hollande apparemn	nant de vous rAejouir e	t suitout allez votei jus	qu A la uerni	IA TE SECONDE. #HON	anue				
7					l'espă"re que Frană	δois Hollande a au moir	s pensã© ã inviter l'a	itre salone de	a Diallo					
8				Franããois Holland	e nourrait mãªme a	tteindre les 54%: IRIB -	Selon les sources de 20	minutes ch n	rÃ@sentes au OG du	P http://t.co/saOpw	Yve			
9				RT @ksnavarra: N	ew noll #France201	2 #TNS-Sofres #Holland	le 55% #Sarkozy Å 45%	6 http://t.co.	/ch3T3P2C #RadioLo	ndres #nrÃ@sidentielle	iy.			
10		NT (WASHADDRIG) A THE AND A THE ATTRACT AND A THE AND A THE ATTRACT ATTRACT AND A THE ATTRACT ATTRACT AND A THE ATTRACT ATTRACT AND A THE ATTRACT ATTR												
11				RT @ZaieMan: Ho	llande fait gaffe pas	d'ouverture sur le mod	em sans le FDG sinon Ã	§a va barder!						
12				Je voulais demand	er a mon pÃ"re pou	r mercredi mais si hollar	nde sort je pense que je	vais m'en pre	endre une					
13				EN DIRECT A #Bas	tille et à #Concord	e, les soirées se pré	parent. Les derniÃ"res i	nfos ici >8	gt; http://t.co/dXNf	DDHf #Hollande #Sarko	zy			
14				RT @Apleonexe: S	ofres : 55 Hollande	- 45 Sarkozy								
15				RT @bembelly: Ici	#RadioLondres \nU	Ininstalling Sarkozy in pr	ogress 99% done \nâ-^	a-^a-^a-^a-	^a_^a_^a_^a_^aaa_	^â–^â–^â–^â–'â–' Runni	ng time: 5 years - Re	maining Time: 4	100â€	
16	43.544	08195	1.342912	@Benoit_Martine	z Hollande. Désol	ée mais je peux pas m	ettre en favori un twee	t avec le non	n du Flanby.					
17				RT @benedictbrog	gan: #Hollande win i	now certain. Why France	e fears what might hap	oen next - my	piece yesterday http	p://t.co/Fjfobnai				
18				"Las primeras teno	lencias dan a FranÃ	§ois Hollande en cabeza	" titula la radiotelevisiÂ	⁽³ n pÃ⁰blica f	rancófona de Bél	gica. 53%-47%				
19				"@elpuntavui: Hol	lande guanya el due	el amb Sarkozy http://t.o	:o/9rvML4vX"							
20				RT @AmelieFish: N	Martine Aubry est ar	rivée au QG de Franco	ois #Hollande pour la "r	éunion stra	atégique" sur les rÃ	©sultats #PRESIDENTIE	LLE ht			
21				RT @duchp: Bien o	des indices annonce	nt un nouveau préside	ent en France : Franço	is Hollande						
22				RT @Benoit_Mart	inez: Si Sarkozy gag	ne je vais démonté l	es tl des gauchistes por	ur faire chier	. Sauf qu'il passera p	as				
23				RT @DSK:.	J'espère que FranÃ	§ois Hollande a au moir	ıs pensé à inviter l'aι	itre salope de	e Diallo.					
24				With a #Hollande	victory in France, La	bour's economic and an	ti-aspiration proposals	are set to be	shown as the poten	tial failures they are.				
25				RT @lesoir: #Holla	ande donnĩ vaing	ueur avec 53 % des voix,	rassemblement UMP a	annulĩ plac	e de la Concorde - ht	ttp://t.co/IPFI4AbO #6n	nai			
26				RT @Salieri_: İlk	sonuħlara gĶre I	Iollande, Sarkozy'nin Ķ	nļndeymiş.							
27				@afterdark175 Og	gnuno ha le sue croc	i: gli elettori di Sarkozy l	han Carla Bruni, i fan di	Springsteen h	han Patti Scialfa. Chit	tarre straziate.				
28				RT @manubenas:	EN DIRECTO: Los m	edios belgas dan por ve	ncedor a Hollande http	://t.co/WMH	Ix5pH4					
29				Frankreich: Hollan	de laut Prognose kl	ar vor Sarkozy http://t.c	o/SJhjND98							
	Sheet1	% Unpopulated	Time histogra	am Create ma	р (+)				4					Þ
Select destina	ation and press E	NTER or choose Paste									Count: 52916	E 🗉		+ 100%
	-													

From the first exercise you'll know that very few records (how many?) are geocoded. To map them select the whole sheet (by clicking on the little arrow top right hand side near A1) then hold down CTRL while unselecting the first row with the column names. Now, in the Home ribbon, click Sort & Filter on the right hand side and select Smallest to Largest. This will put the non-NULL records together at the top of the sheet.

Use the mouse to select the cells that have Latitude and Longitudes along with the text of the Tweet and hit CTRL+C to copy them. You should go from row 1 to row 737 as below:

Aı	toSave 💽 Off	ت ا	· @ - •			fred - Excel			Adrian	Tear 🗈	- 0	\times
Fik	Home	Insert	Page Layout Formulas Da	ıta Review View Help ♀⊺e	II me what you want to do							Share
Paste	X Cut Copy → Format Pai Clipboard	nter	bri <u>•</u> 11 • A Ă Ă <i>I</i> <u>U</u> • ⊞ • <u>○</u> • <u>A</u> • Font □	= = ≥ ≫ - PWrap Text = = = € 5 Harge & Center Alignment	General ▼ ▼ S ~ % * 50 00 00 00 00 00 00 00 00 00 00 00 00	Conditional Format as Formatting + Table +	lormal Bad leutral Calculation Styles	Good · · · · · · · · · · · · · · · · · ·	Insert Delete Format Cells	∑ AutoSum ↓ Fill ~ Clear ~</td <td>Sort & Find & Filter - Select</td> <td>~</td>	Sort & Find & Filter - Select	~
	j n	*	×	fred		d) by l	hitting CTRI +DOWN ARROW; i	n the blank cell I see the value 0.				
	А		В						С			
719	5	2.4949565	5 -1.8812685	Turnout in second round of French pr	esidential election was 8	1.5%						
720	5	2.9537615	-1.166504	Lots of people calling it already for H	ollande @fhollande #FH2	012						
721	53	07530022	11.85859966	Hollande!! #fr2012								
722	53	27350265	-6.39952115	Looking really good for #Hollande at	n! http://t.co/3seV15Ls							
723	53	33971353	-6.26176549	@Brodinski Facile de gagner autant d	e tune a l'étranger et d	e la ramener aprÃ"s pro he	ollande					
724	53	34331265	-9.10639065	Looks like Hollande has won. Does th	is mean journalists will tr	eat us to Clouseau type pr	ronunciations of his name for n	ext few years? O-land				
725	53	35490235	-6.24740065	@MelissaBellF24 reports from #Sarke	zy's election night headq	uarters in #Paris http://t.o	co/bjXMb4Ie #france2012					
726	53	35490235	-6.24740065	Hollande est annoncé gagnant par	es #belges. J'espère que	tu as mauvaise conscienc	ce @Anissamayy					
727		53.403907	7 -1.563097	a very impressive turnout in the Frend	h electionsnow lets see	e the right result! #Holland	de					
728	5	3.8224415	-1.545356	Looks like Hollande has done it! Nice	one!! #VoteLeft #France							
729	5	3.9092965	-9.4202576	I quite like Hollande.								
730		54.535247	7 -1.5585465	@tomilo My poor French knowledge	made me think you were	saying Arrét Hollande th	here. However, I agree!					
731		54.535247	7 -1.5585465	Wonder just how much Franã§ois Ho	llande has won by.							
732	54	57673502	-5.98026252	"@afneil: Leading Brussels paper	eporting Socialists claimi	ng Hollande has won 53%	o to 47%. #France2012†\nOh j	joy! Au revoir sarkozy!				
733		55.019288	-1.652436	Can someone explain me the pro's an	d cons of #Hollande winr	ning the elections? #france	e #sarkozy					
734	57	69391366	5 11.9554135	Change or no change? #hollande								
735	5	9.2833976	5 18.08178341	Sarkozy ut, Hollande in. Fall framåt.								
736		59.333671	L 17.9800231	Det finns sĤkert en massa man kan t	ycka om Hollande, men a	itt socialisterna vinner pre	esidentposten i Frankrike Ĥr br	a för vänstern i Europa.				
737	5	9.9583986	5 17.7766442	Hollande ser ut att vinna det franska	oresidentvalet! #eupol\nł	http://t.co/HHUrIJxx						
738		C	0 0									
739				http://t.co/5McIYPxW Sarkozy, Holla	nde : qui sera le prochain	président ?: La participa	ation progresse par rapport h	http://t.co/h8AtW7mw				
740				Els primers sondatges atorguen la vic	:òria a Hollande - VilaW∉	eb http://t.co/bxehT76W	vÃ-a @vilaweb					
741				Mais j'ai forcemment hate de voir ho	lande a l'oeuvre x) *Ame	n*						
742				RT @nurtapa: Attendez le résultat	avant de vous réjouir et	surtout allez voter jusqu'	'Ã la derniÃ"re seconde. #holla	nde				
743				La France devient Hollande apparemi	nent.							
744				RT @DSK: J'espÃ"re que Frank	§ois Hollande a au moin:	s pensé à inviter l'autre	salope de Diallo.					
745				François Hollande pourrait même a	tteindre les 54%: IRIB - S	elon les sources de 20min	nutes.ch présentes au QG du I	P http://t.co/saQpwnyX				
746				RT @ksnavarra: New poll #France201	2 #TNS-Sofres #Hollande	e 55% #Sarkozy à 45%. ht	ttp://t.co/cb3T3P2C #RadioLor	ndres #présidentielle				
747				RT @NicolasBedos1: -Tu vois vraime	nt une différence entre	Hollande et Sarko? - Oui,	Marine Le Pen.					
	Shee	t1 % U	Inpopulated Time histogr	ram Create map +								•
Ready							Average: 18.49	928017 Count: 2211 Sum: 272	30.94041) 🗉 – —		► 100%

In the Chrome web browser go to:

https://batchgeo.com/

BatchGeo: Create an interactive map from your data - Google Chrome									
← → C Secure H	https://batchgeo.com						· · · · · · · · · · · · · · · · · · ·	☆ 🔤 🗄	
							pro sign-in	,	
	hatch	000		make mana feature		vising	hilo ournert bles		
		igeo		make maps teatur	es pro p	fricing mor	blie support blog		
		•							
		Paste vo	ur lo	cation data	below	to map	it [.]		
	'			oution data					
	Example Address	City	State Zip	Name	Phone Number	Group	URL		
	1 Crossgates Mall Road	Albany	NY 122	3 Apple Store Cross Gates	(518) 869-3192	Example Group 1	http://www.apple.com/retail/		
	Duke Rd & Walden Ave	Buffalo	NY 142	25 Apple Store Walden Galleria	(716) 685-2762	Example Group 2	http://www.apple.com/retail/		
	630 Old Country Rd.	Garden City	NY 115	Apple Store Roosevelt Field	(516) 248-3347	Example Group 3	http://www.apple.com/retail/r		
	160 Walt Whitman Rd.	Huntington Station	NY 117	6 Apple Store Walt Whitman	(631) 425-1563	Example Group 3	http://www.apple.com/retail/		
	(Don't forget to inc	lude some header colun	nns - You can Vali	also try our Spreadsheet Template	e (Excel), or hit "Ma Map Now	p Now" and try it out	with our example data.)		
	Create a map:								
	Address City State 2 Crossgates Mail Roac Albany NY Duke Rd & Walden Ave Buffalo NY	D E F Zipcode Name Phone I 12203 Apple Store Cros (518) 86 14225 Apple Store Walc (716) 65	Bosic e	Region United States 8 Location / Address Example Address 8	Marker box preview Apple Store Cross Gates 1 Crossgates Mail Road Alberry, NY 12203		ge.commen ⁻ titelecko ² /filmeseidedbadts		
	4 630 Old Country Rd. Garden City NY	11530 Apple Store Root (516) 24		City / County City 0	Phone Number: (518) 869-3 Group: Example Group 1	 A 	@San togeto		

Paste the cells you have copied from Excel into the obvious box (it goes green outlined when you click on it)...

Once it's pasted in you should see something like the below:

BatchGeo: Create an interactive map from your data - Google Chrome										
C 🔒 Secure https://batchgeo.com		ι () ()				☆ 🛽 🚥 🗄				
					pro sign-in 🔹					
o batchgeo	make maps fea	atures pro pricin	g mobile	support	blog					
Paste you	ur location data	a below to	map it:							
interaction_geo_latitude	interaction_geo_longitude	interactio	n_content							
-36.73745265	-71.03529455	Vaamos S	arkozyIIIII :(
-36.73745265	-71.03529455	Qué loc	o que segÃ⁰n sondeos	s - belgas y suizos	s - gan					
-36.73745265	-71.03529455	Cagê Sa	kozy????\nShuuuuat	a y la Carlita?						
-36.73745265	-71.03529455	@bdelam	aza serÃj 47% - 48% i	de Sarkozy??						
(Don't forget to include some header column	ns - You can also try our Spreadsheet Tem	plate (Excel), or hit "Map Nov	" and try it out with ou	ir example data.)						
Create a map:										
Address City State Zocolo Name Phone I Address City State Zocolo Name Phone I Consignities Mail Road-Abany NY 12020 Apple Store Xocol (518) 6f Duale R& Marden Are Suffalo NY 14220 Apple Store Work (718) 6f 4 600 Clocolny R6 (3 dearh Cit NY 1550 Apple Store Root (161 2)	Basic optime Pegion Loostion / Addess City / County Oty	Marker los preview Apple Store Cross Cates Cocegoios Mal Road Alberg, NY 12203 Phone Number, 0518) 800-0 Groups Example Group 1		traticul 75.6 teder 446 303.4 0e		·				

Next click on the button Validate & Set Options...

BatchGeo: Create an interactive map from your data - Google Chrome										
← → C			☆ 🖾 🗉 :							
Validate & Set Options We have made our best guess at your data, but check below that everything is ok.										
Basic options Marker box preview										
Region	International •	interaction_content: Vaamos Sarkozy!!!!!								
Location / Address	none •	:(
City / County	none •									
State / Province / Postcode	none 🔻									
Zip / Postcode / Country	none 🔻									
Group By / Thematic Value	Single Color •									
	Show Advar	Map								
Create a map: 1 defres City State Speech Parts 2 lossagete Mal Rock-Anno Burldo 3 Date R4 StateIn Anno Burldo 4 00 OH County R4 careful Anno Parts 5 100 Wall Whitmain R4. Hurtington StVT 11226 Apple Borne City (1) 5 100 Wall Whitmain R4. Hurtington StVT 11764 Apple Borne City (2) 5 100 Wall Whitmain R4. Hurtington StVT 11764 Apple Borne City (2) 5 2065 Rodmond Ave: Statem Island WT 10314 Apple Borne Barl(7)(4)	Tear system United Sectors (CP) Carry Or Star Product Frankis Dr. Phatos Frankis Dr. Phatos Carrier Dr. Phatos View Orac	Minute Interpretation Minute Interpretation Alterna (N 1020) Compared Autor (N 1020) Compare								

Then click the further button Show Advanced Options...

© Batch	p from your data - Google Chrome	_ @ ×	
	Geo: Create an interactive main All Columns • Use Google Maps • interaction_geo_latitude • interaction_geo_longitud •	p from your data - Google Chrome	
	Mak	te Map	
Create a map:			

It should have correctly picked out the Latitude and Longitude columns. Adjust the Title column to use the field **interaction_content** as below:

© Batc	hGeo: Create an interactive ma	p from your data - Google Chrome	_ ð ×
← → C			☆
Advanced options			•
Title	interaction_content •	Label each marker none	
Marker Description	All Columns 🔹	Select a default map view to show street	
Country Field	•	Enable clustering for high density markers	
URL	Use Google Maps 🔹	Calculate (straight line) distance from first address in miles	
Image URL		Hide map addresses / Limit zoom in	
E-mail	•	Disable KML export	
Latitude	interaction_geo_latitude •	 Links open a new Window 	
Longitude	interaction_geo_longitud	Troubleshooting	
		Filter road types before geocoding	
Marker colors	Marker shape	Map style	
Color Group name			
Marker			
		A Contraction of the second se	
		Low Broakyn	
	Mak	ке Мар	
	_		
Create a map:			
			· · · · · · · · · · · · · · · · · · ·

Finally, click the Make Map button to plot your geolocated Tweets...



You will see where most geolocated Tweets are concentrated world-wide. Zoom in on Paris (or wherever you like) to click on individual Tweets and see what the user is saying...



This processing has used Excel and BatchGeo to produce the map. You can save this using the Save & Continue button if you wish; please **do not Share** the map, **keep it Unlisted** and email yourself the link...

BatchGeo: Create an interactive map from your data - Google Chrome								
← → C ▲ Secure https://batchgeo.com		☆ 💷 😳						
	Make Map Geocoded 756 records.							
Title F	French Election Tweets							
Provide Description	interashortule for your map, what is it? Some <u>mappable</u> Tweets from the French Presidential Election 2012							
Montessor 2000 Email e	email@myport.ac.uk							
Share Construction	Public Unlisted Password Protect (Pro Only) fyou select "Unlisted" your map will be saved to a unique URL that will be kept vivate to you, it will not be listed anywhere on the site. You can read up on map							
	anvacy in our readures section. ☑ I have read and agree to the BatchGeo Terms of Service.							
Colle Saint Conti	Save Map							
Google (Microson) 550	× Versions VIII							
		· · · · · · · · · · · · · · · · · · ·						

The Department uses Geographic Information System (GIS) software to perform these sorts of operations on, potentially, much larger data sets. Data may also be saved in the Relational Database Management System (RDBMS) called Oracle – on in other relational databases – for query and analysis. Ask one of the supervisors for more detail if you're interested or consider the GIS module as an option for your Degree programme...

What else?

Even though this spreadsheet only contains a subset of the many fields/columns saved when downloading data from Twitter (see the file **French Social Media Data - All Data (for those interested).xlsx** if you're interested) there are still a few other things you can easily achieve with some copying/pasting and free online tools...

Wordclouds



Copy some of the interaction_content cells and paste them into https://www.jasondavies.com/wordcloud/

Other word cloud generators online let you upload files. You can try finding another one (e.g. <u>http://www.wordle.net/create</u> or <u>https://tagcrowd.com/</u>) and copy/paste more cells from the **interaction_content** column into a text file (use Notepad), saving this as WORDS.TXT and try uploading this to analyse all of the words. Note that some sites (e.g. Tagcrowd) have a 5MB limit of file size so you may not be able to create a word cloud visualisation of *all* the records in the spreadsheet.

Question: What do word clouds help you achieve? Does the relevant visual weight of the words identified in the text help you draw any conclusions about what people were Tweeting at the time the results came in during the 2012 French Presidential Election? Who, somewhat surprisingly, won that election?

Ranking

There are several fields in the Excel spreadsheet (e.g. klout_score, twitter_user_followers_count, twitter_retweet_user_followers_count and twitter_retweet_count) that can be used to rank (or sort) data. Highlight all cells in the spreadsheet (little arrow near cell A1) then from Home ribbon hit Sort & Filter and Custom Sort... Select one or more levels of sortation (e.g. on klout_score largest to smallest as above) and look at the Tweeters at the top of this list:

		? ×			
Level XDelete Lev	el	🖹 Copy Level 🔺 🔻	Options	✓ My d	ata has <u>h</u> eaders
		Sort On		Order	
klout_score	\checkmark	Cell Values	~	Largest to Smallest	\checkmark
				ОК	Cancel
	Level Delete Lev	Level <u>Pelete Level</u> klout_score	Sort I Level Delete Level Copy Level Sort On klout_score Cell Values	Level € Copy Level Options Sort On klout_score Cell Values	Level ©ptions Sort On Order Klout_score Cell Values OK

AutoSave 💽 Off) 🔚 <table-cell-rows> ଟ - =</table-cell-rows>		fred - Excel		Adrian Tear 🛛 🚽 🗇 🗙
File Home Insert Page Layout Formulas Data Review View He	Ip ♀ ♀ Tell me what you want to d			函 Share
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	e & Center - Son		Good · · · · · · · · · · · · · · · · · ·	Insert Delete Format Cells ∑ AutoSum A ∑ Fill Sort & Find & Filter - Select - Editing
fre	d	 by hitting CTRI +DOWN ARROW; 	in the blank cell	
		, our be time to see to go you should	d see the value 0.	
D	E	F	G	H
1 interaction_source	interaction_author_username	interaction_author_name	demographic_gender	klout_score interaction_content
2 (0 0	C	0	0
3 WSJ.com	WSJ	Wall Street Journal		87 Breaking: François Hollande appears to h
4 HootSuite	el_pais	EL PAIS	mostly_female	86 RT @elpais_inter: Pase lo que pase hoy en
5 HootSuite	el_pais	EL PAIS	mostly_female	86 Los primeros sondeos dan a Hollande la pr
6 bitly	la_patilla	La Patilla		85 Sondeos afirman que Hollande le ganA ³ a S
7 TweetDeck	Reuters	Reuters Top News		84 Follow 2012 French Election coverage here
8 web	agarzon	agarzon		82 La posible victoria de Hollande puede ser d
9 web	NoticiasCaracol	Noticias Caracol		82 FranA§ois Hollande serA-a el nuevo preside
10 breakingnews.com	BreakingNews	Breaking News		81 FranA§ois Hollande appears to have won t
11 TweetDeck	Milenio	Milenio.com		79 Se enfrentan Sarkozy y Hollande en segund
12 TweetDeck	noticias24	noticias24		79 Francia: Exit polls dan la victoria a Holland
13 web	repubblicait	la Repubblica		79 La non-stop su @repubblicait: Sarkozy ann
14 TweetDeck	noticias24	noticias24		79 SegÃon los Exit Polls Hollande es el nuevo p
15 HootSuite	lemondefr	Le Monde	mostly_female	79 Sarkozy aurait décidé d'avancer son al
16 VEJA	VEJA	VEJA		78 Pesquisas indicam a vitÃ ³ ria de Hollande co
17 Twitter for BlackBerry®	AJEnglish	Al Jazeera English	mostly_male	78 RT @AJELive: AFP: that according to estim
18 web	elmundoes	elmundo.es		78 Los sondeos dan ganador de las elecciones
19 web	cuneytozdemir	cļneyt Ķzdemir	male	78 Fransa'da seħimleri Sarkozy kaybetmiğ.
20 Cooperativa.cl	Cooperativa	Cooperativa.cl		78 Sondeos suizos y belgas dan triunfo a Holla
21 HootSuite	JornalOGlobo	Jornal O Globo		77 Pesquisas de boca de urna indicam vitÃ ³ ria
22 TweetDeck	CNBC	CNBC		77 WSJ: FranA§ois Hollande appears to have v
23 Tweet Button	rtve	RTVE		77 En vivo electoral: Los primeros sondeos da
24 TweetDeck	rtve	RTVE		77 Minuto a minuto electoral: Los primeros se
25 web	Chicagokrzy	Jim Minardi	male	77 RT @MichelleFields: Hollande Defeats Sark
26 TweetDeck	BBCWorld	BBC News (World)		77 Confident mood outside Socialist HQ in Par
27 TweetDeck	BBCWorld	BBC News (World)		77 Sense of defiant, determined, almost despe
28 TweetDeck	doonie	Dino Amenduni	male	76 II giornale belga Le Soir: Sarkozy avrebbe a
29 Facebook	doonie	Dino Amenduni	male	76 A questo punto si puA ² dire: Francois Holla
Sheet1 % Unpopulated Time histogram Create map -	÷	•		
Ready				Count: 662977 🔠 🗉 🖳+ 100%

You should see that the top Tweeters, in terms of **klout_score** (see <u>https://klout.com/corp/score</u> for further details), are large news organisations like the *Wall Street Journal* and *El Pais*.

Question: Who are the least important Tweeters by **klout_score**? Is the average **klout_score** for those Tweeting with geographic coordinates higher or lower than for those who Tweet without coordinates?

Tip: Use Excel to calculate averages and make these comparisons using what you have learnt above.

Sentiment analysis

The spreadsheet also contains a column **salience_content_sentiment** which gives an indication of negative, neutral or positive sentiment for the Tweet.

Select this column and paste it into a new sheet...



From the Insert ribbon choose Pivot Table and follow the defaults to create another new sheet:

	Create PivotTable	? ×											
Choose the data that you want to analyze													
<u>S</u> elect a table or range													
<u>T</u> able/Range:	1												
O <u>U</u> se an external data source													
Choose Connection													
Connection nar	ne:												
O Use this workbook's	Data Model												
Choose where you want th	ne PivotTable report to be placed												
• <u>N</u> ew Worksheet													
O <u>E</u> xisting Worksheet													
Location:		1											
Choose whether you want	to analyze multiple tables												
Add this data to the	Data <u>M</u> odel												
	ОК	Cancel											

This will create a blank definition:

	AutoSave	• Off		ي ∙ ر	- =		_			PivotTab	le Tools		fred - Exce	el						Adrian	Tear 🗇 -	- a ×
		Home	Inser	t Page I	ayout Fe	ormulas Da	ata Re	eview View	Help	Analyze	Design	🖓 Tell me	what you wa	ant to do								년 Share
Pive Piv	otTable Na otTable11 Options PivotTable	ame: Ac	tive Fi	eld: Settings	Drill Drill Down Up ~ Active Field	*∃ Expand F =∃ Collapse	Field	→ Group Select ○ Group Field Group	Inser Slice	t Insert r Timeline Filter	Filter Connections	Refresh C	hange Data Source + Data	Clear Se	elect Move • PivotTable	Fields, Items, OI & Sets - Too Calco	f _x AP Relationshi ols -	ips PivotCl	art Recomme PivotTat Tools	nded Field bles List	d +/- Field Buttons Headers Show	~
		h		×					fred					BU F	ov hitting CTRI	+DOWN ARROW	'; in the blank uld see the val	cell ue 0.				
	A	1	в	с	D	E	F	G	н	1	J	К	L	м	N	0 P		R	S	T A		
1						-							-				-		-		PivotTabl	e F 🍷 🛛 ×
2 3																					Choose fields to report:	add to
4		Pivot	Table1	1																	Search	Q
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	To built from t	d a repo	ert, chu tTable	oose fields Field List																	☐ salience_con More Tables Drag fields bett ▼ Filters	tent_sentiment ween areas below:
21																						
22																						
24																					≡ Rows	Σ Values
25 26 27 28																						
29		Shee	et1 🗆	% Unpor	ulated	Time histor	ram 🖂	Create man	Sheet15	Sentin	nent i	A									Defer Layou	t Upda Update
Rea	ady	June										0									I II	+ 100%

On the right hand side drag salience_content_sentiment into the Rows and Sum Values boxes:



Then from the Insert ribbon choose Insert Recommended Chart:



This should suggest a Clustered Column chart as below:

AutoSav	AutoSave 🖲 Off) 🖬 🕤 - 🔿 - =								PivotChart Tools																						
File	Home Insert Page Layout Formulas Data Review View Help								Desig	n Fo	ormat	Q Tell me what you want to do																			
Paste	Cut Calibri (Body) • 10 • A A = = = \gg · P Wrap Tex Copy • B I U • = · Δ • E = \equiv = $=$ = \equiv = \equiv = $=$ = \equiv = $=$ = = $=$ = = = $=$ = $=$ = $=$ = = = $=$ = $=$ = $=$ = $=$ = $=$ = $=$ = $=$ = $=$ = $=$ = $=$ = = = $=$ = $=$ = = = $=$ = = = $=$ = = = $=$					t Center ~	General	9 4	v 00. 0, 0.€ 00	Condit	Conditional Formatia - Table - Normal Bad Good									Ins	ert Dele	ete Form	nat	∑ AutoSum ~ A Z ♥ ↓ Fill ~ ◆ Clear ~ Sort & Find & Filter ~ Select ~							
Clipb	oard r	5 Font	F	2		Alignme	nt		Nur	mber				3)	by hitt	ting CT	Styl	les OWN A	RROW	: in th	ne blan	k cell			Cel	lls		Editing		^	
	n	- ×					fred	an overting CTRE+DOWN ARROW; In the blank cell																							
	Α	В		C	DE	F	G	4 I I	1.1	K	1	м	N	0	0 0	R	S	Т	ίπ.	v	w	x	·	7		BACI					
1	A	0		C	U		0 1		,	ĸ		IWI	1.4	0	Q	K	5		0	•	~~	~		~		U AC 7		PivotChart	F 👻	×	
2																												Choose fields to a	dd to		
3 Row La	abels 💌 Sun	n of salience_conter	nt_sentimen	t																								report:		<u>Ω</u> τ	
4 -15			-1	15																								Search		0	
5 -14			-5	56																										~	
6 -12			-6	50																								✓ salience_conte	ent_sentime	ent	
7 -11			-2	22		0				0						0															
8 -10			-504	10		Sum of	salience_con	ent_sentim	ent	-						F															
10 -9			-3873	54		-				Tetal																					
11 -7			-360	25		_				Iotai																					
12 -6			-357	76		30000																									
13 -5			-445	50		20000						-																			
14 -4			-563	36		10000						1.1																			
15 -3			-720	00		0			uue:	1.00	d de					0															
16 -2			-230	8		-10000	\$ \$ \$,¢ ,o	2	0 2	D 6	\$ \$ \$	5 2	~5	Total	i T															
17 -1			-120)5		-20000																						Drag fields betwe	en areas be	elow:	
18 0				0		-30000																						T Filters	III Legend	d (Se	
19 1			130	00		-40000																									
20 2			181	10		-50000																									
21 3			342	23		calioner	content cor	timont w																							
22 4			560	00		Salience	s_content_ser	ument +																							
23 5			327	/5		Ŭ																						= Avic (Cate	Σ. Values		
24 6			450	00																								- Axis (Cate	Z values		
25 /			703	20																								salience *	Sum of sa	al *	
27 9			2263	25																											
28 10			207	70																											
29 11			4	14																											
	Sheet1	% Unpopulated	Time hist	ogram	Cr	eate map	Sentim	ent Pi	rvot Tabl	le C	ondit	ion	+	4	-			-						-		Þ	•	Defer Layout	Jpda Up	pdate	
Ready																										Ħ	=	<u> </u>	+	+ 100%	

Question: What does this data tell us? It appears that large numbers of Tweets are Negative (-9) but that quite a few (+9) are Positive. Is this meaningful?

Tip: Use Conditional Formatting / Colour Scale to colour the spreadsheet according to **salience_content_sentiment** by copying this column, and **interaction_content** into another new sheet:

Aut	oSave 💽 🔐 🔓 🗧 🔶 -	÷				fred - Excel					Adrian	Tear 🗈	— c	J X
File	Home Insert Page Lay	yout Formulas Data	a Review V	iew Help ♀Te	II me what you want to do									년 Share
Paste	Cut Calibri Copy → Format Painter Clipboard	• 11 • A A = • 1 A A = Font 5	E = = ≫ • F = = € → Align	Wrap Text	General → → ♀ * % * 00 *00 *00	Conditional Format a Formatting - Table -	Normal Neutral	Bad Calculation Styles +DOWN ARROW: in	Good Check Cell	Insert Dele	te Format	∑ AutoSum ↓ Fill ~ Clear ~ E	Sort & Fin Filter * Se	nd & lect *
	n 👻 🗙			fred		-		, ou should	see the value 0.					
	A								B					
28	ll gi	giornale belga Le Soir: Sa	arkozy avrebbe a	nnullato la festa in p	lace de la Concorde (ho	me di Repubblica)								
29	Ag	questo punto si puÃ ² dir	e: Francois Holla	nde.		ne arriepassiea,								
30	7 Hol	ollande on course for vic	ctory: French eler	tion live with @mat	tholehouse http://t.co/-	19gP5X5k								
31	0 500	cialist candidate Franco	ois Hollande has v	von the French presi	dency, according to esti	nates: http://t.co/49g	P5X5k							
32	0 Hol	ollande gana a Sarkozy, s	según los sonde	os publicados en Sui	za http://t.co/0DQb7Vil	#France2012								
33	8 Los	os medios belgas y suizos	s avanzan una vic	toria de Hollande ht	tp://t.co/PgTWXwoW									
34	5 Est	timates show Hollande	winning French e	lection: WSJ http://t	.co/0zUfDDht									
35	Alle	lez Sarkozy-Hollande htt	tp://t.co/UKWGo	ubL #Sarkozy #Holl	ande									
36	9 RT	i @rtve: En vivo electora	al: Los primeros s	sondeos dan la victo	ria a Hollande con el 539	6 de los votos http://t	.co/5y8LAwAs v	Ă-a @rtve						
37	1 Sol	bre Francia, los sondeos	s dan la victoria a	Hollande. En Grecia	a los conservadores de	Nueva Democracia. D	e momento sóle	o sondeos.						
38	2 Los	os exit polls dan ganador	r a Hollande.											
39	-4 The	e first rule of French ele	ection results figh	it club is you don't ta	lk about French electior	results fight club								
40	9 "@	⊉acentodiario: La prensa	a belga da como	ganador a Hollande	con el 53 % de los votos	http://t.co/rCJ0Tqgx	#PorUnMejorPai	sVota1 @llegopapa						
41	0 Fol	llow @Reuters 2012 Fre	ench Election cov	verage here: http://t.	co/8JGZHNuK									
42	0 Pes	squisas apontam vitÃ ³ ri	ia de Hollande na	França. http://t.co	/CHcRjhAX									
43	3 Sar	rkozy adelanta su declar	ración prevista p	oara las 21.30 a las 2	0.00vÃ-a @elpais_int	erya debe tener to	do claro							
44	0 RT	@afneil: Leading Bruss	els paper reporti	ng Socialists claiming	g Hollande has won 53%	to 47%. #France2012								
45	#Fr	rance #Sarkozy #Holland	de «سارÙfi	Ù^زي» Ù^«هÙ	^لانØ⁻» ÙŠØ⁻Ù"I	IJŠØ§Ù†Ø"ØμÙ^Ø≇ÙŠ	ù‡ùø§ ùùš ø	¥Ø¹Ø§Ø⁻Ø© ان¢	ðªØ®Ø§Ø¨Ø§Øª ال	رئاسة	ø§ù"ùø	±Ù†Ø³ÙŠØ©	http://t.co	o/nBTxdl2
46	-6 @F	Frenpega pero seria el m	nenos peor, a esc	o se le tira en Francia	ı, a la salida de un sarko	y servil a estados unic	los y fascista							
47	0 Me	edios europeos adelanta	an triunfo de Hol	lande sobre Sarkozy	en elecciones de Franci	a http://t.co/StYGWAD	Dr vÃ-a @sdpnot	icias						
48	0 RT	∫@paolo_levi: Urgent: №	Nicolas #Sarkozy	a annule' le rassemb	lement populaire préາ	ru ce soir a' Place de la	Concorde. (Sou	rce diplomatiqu						
49	8 Sin	n duda: Hollande abrirÃj	i una brecha fisca	al como la de ZP "	@AinaDiazV: @pedroj_	Crees que Sarkozy ser	Ã-a mejor para E	uropa q Hollande?å	€					
50	3 Too	odas las israelitas difundi	idas fuera de Fra	ncia estÃin dando ve	entaja de 5 puntos a fave	or de Hollande. SerÃj r	nuy difÃ-cil que (eso dé la vuelta.						
51	#Fr	rance2012 A meno di du	ue ore dalla chius	ura dei seggi, #afflue	enza al 70%. #Hollande o	lato al 53% [cc: @stef	_montefiori]http	://t.co/PivohYG8						
52	0 RT	í @capitalsalvaje: @ped	droj_ramirez. Hol	lande darÄj el equilit	orio a Europa que ahora	le falta. Dicen las mala	as lenguas que R	ajoy lo estÃj						
53	2 Est	to es lo que hay. Los sor	ndeos dan ganad	or a Hollande http://	/t.co/ovFu03Fa vĂ-a @e	Imundoes								
54	Car	ırla Bruni Sarkozy dejarA	Al de ser primera	dama. CrA ³ nica de u	n divorcio anunciado									
55	6 Pas	ise lo que pase hoy en Fr	rancia, FranA§ois	Hollande ya ha gana	ado, atirma @BernaGHa	rbour en su blog Puert	as Abiertas http:	//t.co/6lsD4k8h						
56					· · · · ·				#NAME?					
	Time histogram Cr	reate map Sentimer	nt Pirvot Tabl	le Conditional f	ormatting (+)			•						Þ
Ready										Count: 97016) 🗉 🗕 —	-	+ 100%

This should show you how the software has scored each Tweet. Is it accurate? You might need to get your dictionary out if your GCSE French is a bit rusty!!

Try some more advanced software from IBM at <u>https://natural-language-understanding-demo.ng.bluemix.net</u> and type in, or copy/paste in, some sentences of your own or, e.g., some copy from the BBC News website. Check the Keywords and Entities detected in the text. Is this software any better? How could you use it?

Summary

This Exercise has guided you through what can be achieved to analyse and map some Online Social Network (OSN) data, primarily sourced from Twitter.

Quite a lot can be achieved using Excel and some free, online resources. More systematic analysis requires a database for data storage, many more records, and more sophisticated processing using specialised software.

Think about the quality, and use, of this data. Also, are there any ethical problems with using data of this type?

Follow the references for more material covering these areas.

References

- Fuchs, C. (2017). *Social Media: A Critical Introduction* (2nd ed.). SAGE Publications. Retrieved from https://uk.sagepub.com/en-gb/eur/social-media/book250870
- Scott, J. (2017). *Social Network Analysis*. SAGE Publications. Retrieved from https://uk.sagepub.com/engb/eur/social-network-analysis/book249668
- Tear, A. (2014). SQL or NoSQL? Contrasting Approaches to the Storage, Manipulation and Analysis of Spatiotemporal Online Social Network Data. In B. Murgante, S. Misra, A. M. A. C. Rocha, C. Torre, J. G. Rocha, M. I. Falcão, ... O. Gervasi (Eds.), *Computational Science and Its Applications -- ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30 -- July 3, 2014, Proceedings, Part I* (Vol. 8579 LNCS, pp. 221–236). Springer International Publishing. http://doi.org/10.1007/978-3-319-09144-0_16