Sentiment detection

Weiai Xu (Wayne), PhD
Assistant Professor
Department of Communication, UMass-Amherst
Email: weiaixu@umass.edu
curiositybits.cc

Assignments

Comments on your assignment 1 have been posted.

Assignment 2 posted! Refer to the instruction posted on Moodle. Start early!

What is *sentiment detection*?

To automatically identify the types and the strength of emotions in text data

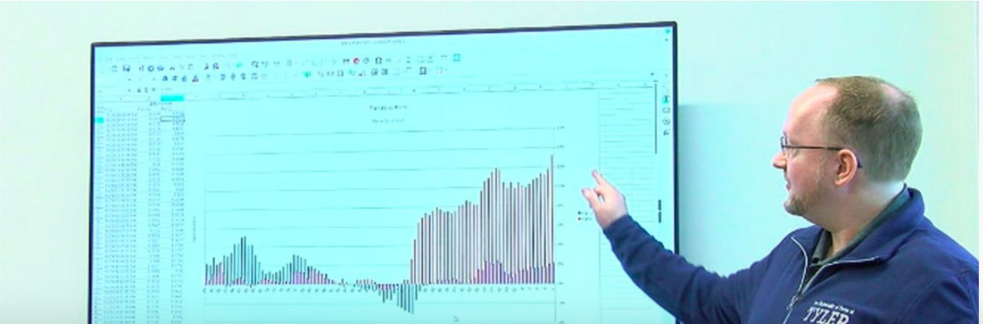
Examples

- Sentiments in tweets, Facebook posts, YouTube comments
- Sentiments in news coverage, speeches, etc.

Examples of *sentiment detection*

UT Tyler professor explains how sports analytics could predict Super Bowl 53 winner

ndeley Desktop



Schumaker looked at data collected through 5 p.m. Tuesday. As of that time, sentiment leaned positively towards the Los Angeles Rams, suggesting that the Rams would win the Super Bowl.

“This also has about 59.8 percent accuracy, so it’s absolutely fascinating to see that Twitter has predictive value in sports,” Schumaker said.

The data analytics lab is a research and teaching lab, where students can learn not only sports analytics, but mine data related to healthcare and pharmatics. Schumaker said the college offers various courses related to analytics.

http://www.kltv.com/2019/01/30/ut-tyler-professor-explains-how-sports-analytics-could-predict-super-bowl-winner/

Examples of *sentiment detection*

What is *sentiment detection*?

Automatic stock trading based on presidential tweets that mention company names

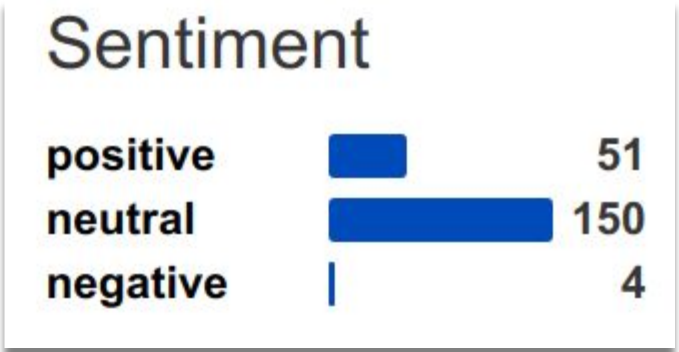
https://www.npr.org/sections/money/2017/04/07/522897876/meet-botus-planet-money-s-stock-trading-twitter-bot

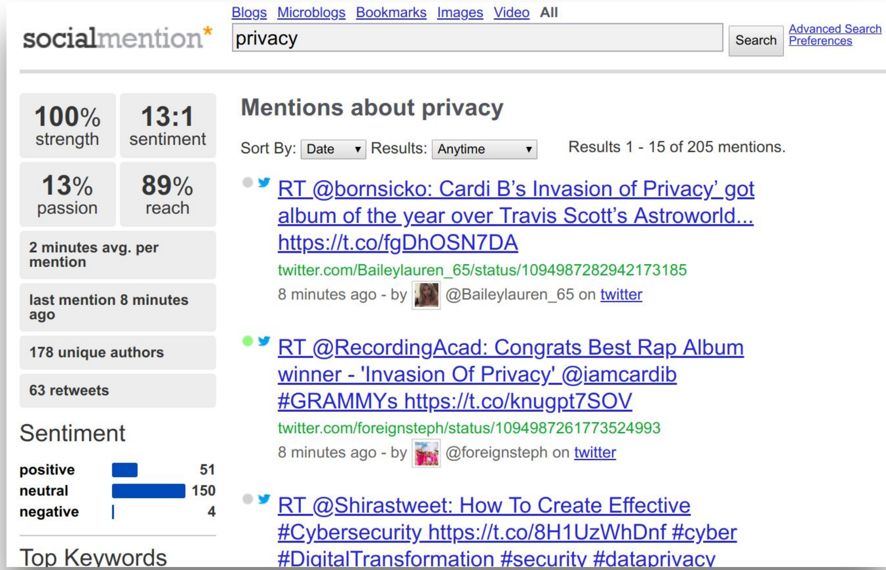
What is *sentiment detection*?

Try socialmention:
http://www.socialmention.com/

Work with your team to go through tweets

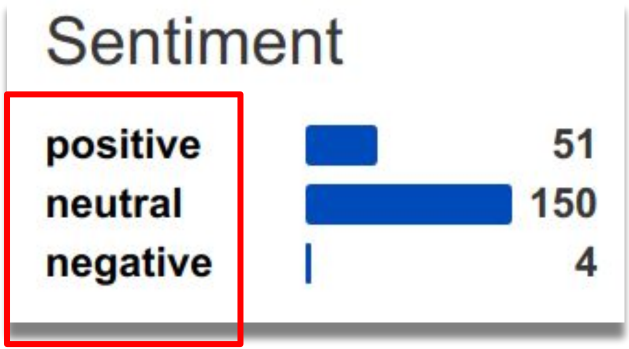
- What types of sentiment?
- Speculate as to why some tweets are marked *positive* or *negative*.





Sentiment types and strength

Sentiment types



The basic classification

Sentiment types and strength

Sentiment types

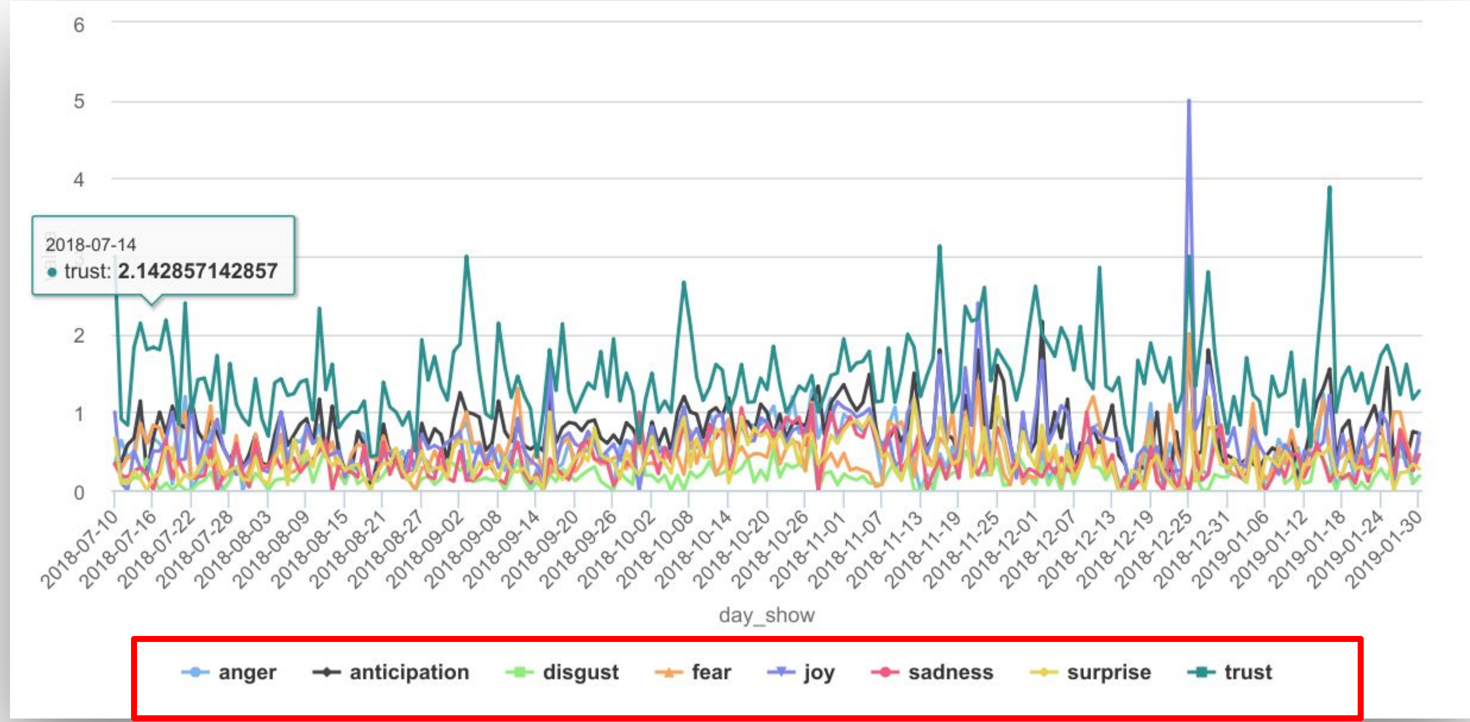
https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

A more complex sentiment detection model.

Sentiment types and strength

Sentiment types

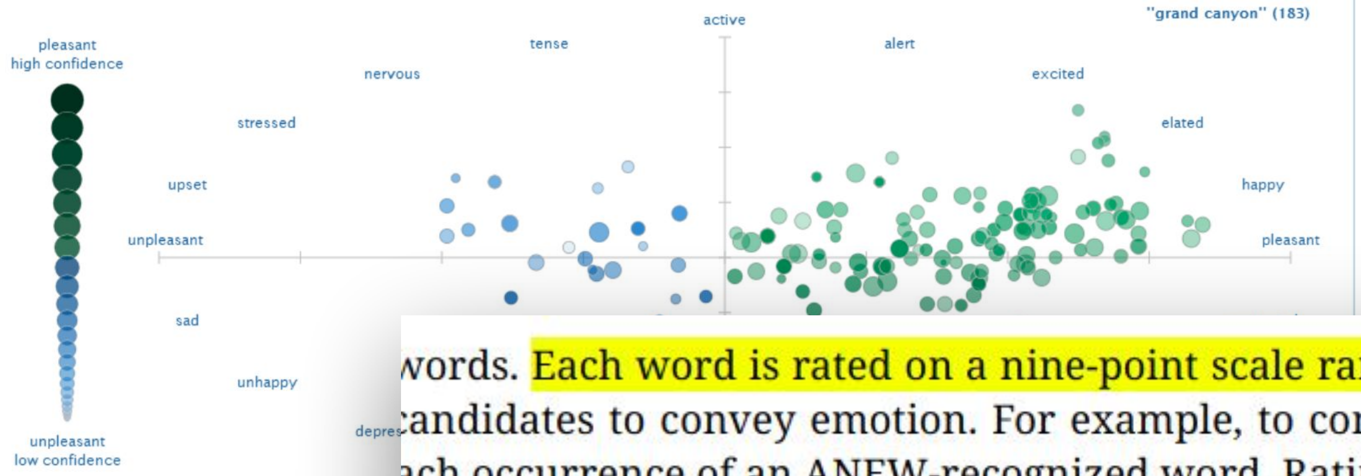
A model we are going to build



Sentiment types and strength

Sentiment strength

A model we are going to build



vords. Each word is rated on a nine-point scale ranging from 1 to 9.
:andidates to convey emotion. For example, to construct the ANEW
ach occurrence of an ANEW-recognized word. Ratings for a common
xample, for the word ***house***, ANEW reports:

Sentiment types and strength

Sentiment strength

Another example

SentiStrength

The text 'I love you but hate the current political climate.' has scale result **-1**.

Approximate classification rationale: I love[3] you but hate[-4] the current political climate .[sentence: 3,-4] [result: max + and - of any sentence][scale result = sum of pos and neg scores] (Detect Sentiment)

Positive sentiment strength ranges from 1 (not positive) to 5 (extremely positive) and negative sentiment strength from -1 (not negative) to -5 (extremely negative). The sentiment strength detection results are not always accurate - they are guesses using a set of rules to identify words and language patterns usually associated with sentiment.

http://sentistrength.wlv.ac.uk/

How are sentiments detected

Dictionary approach

Each sentiment dictionary includes a set of words indicative of sentiment. Algorithms scan texts and calculate the proportion of the texts that contain sentimental words

Some dictionary examples: Happiness dictionary

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0026752#s4

Each word is rated on the scale of happiness

word	happiness_r	happiness_avera	happiness_standa	twitter_rank
laughter	1	8.5	0.9313	3600
happiness	2	8.44	0.9723	1853
love	3	8.42	1.1082	25
happy	4	8.3	0.9949	65
laughed	5	8.26	1.1572	3334
laugh	6	8.22	1.3746	1002
laughing	7	8.2	1.1066	1579
excellent	8	8.18	1.1008	1496
laughs	9	8.18	1.1551	3554
joy	10	8.16	1.0568	988
successful	11	8.16	1.0759	2176
win	12	8.12	1.0812	154
rainbow	13	8.1	0.9949	2726
smile	14	8.1	1.0152	925
won	15	8.1	1.2164	810
pleasure	16	8.08	0.9655	1497
smiled	17	8.08	1.066	--
rainbows	18	8.06	1.3603	--
winning	19	8.04	1.049	1876
celebration	20	8.02	1.5318	3306
enjoyed	21	8.02	1.5318	1530
healthy	22	8.02	1.0593	1393
music	23	8.02	1.1156	132
celebrating	24	8	1.1429	2550
congratulat	25	8	1.6288	2246
weekend	26	8	1.2936	317
celebrate	27	7.98	1.1516	1606
comedy	28	7.98	1.1516	1444
jokes	29	7.98	0.9792	2812

abuse	10190	1.88	1.2395
diseases	10191	1.88	0.9398
sadness	10192	1.88	1.1891
violence	10193	1.86	1.05
cruel	10194	1.84	1.1493
cry	10195	1.84	1.2835
failed	10196	1.84	0.9971
sickness	10197	1.84	1.1843
abused	10198	1.83	1.3101
tortured	10199	1.82	1.4241
fatal	10200	1.8	1.5253
killings	10201	1.8	1.5386
murdered	10202	1.8	1.6288
war	10203	1.8	1.4142
kills	10204	1.78	1.2337
jail	10205	1.76	1.0214
terror	10206	1.76	1.0012
die	10207	1.74	1.192
killing	10208	1.7	1.359
arrested	10209	1.64	1.0053
deaths	10210	1.64	1.1386
raped	10211	1.64	1.4251
torture	10212	1.58	1.0515
died	10213	1.56	1.198
kill	10214	1.56	1.0529
killed	10215	1.56	1.2316
cancer	10216	1.54	1.073
death	10217	1.54	1.2811
murder	10218	1.48	1.015
terrorism	10219	1.48	0.9089
rape	10220	1.44	0.7866
suicide	10221	1.3	0.8391
terrorist	10222	1.3	0.9091

Some dictionary examples: NRC sentiment dictionary

Used in our example

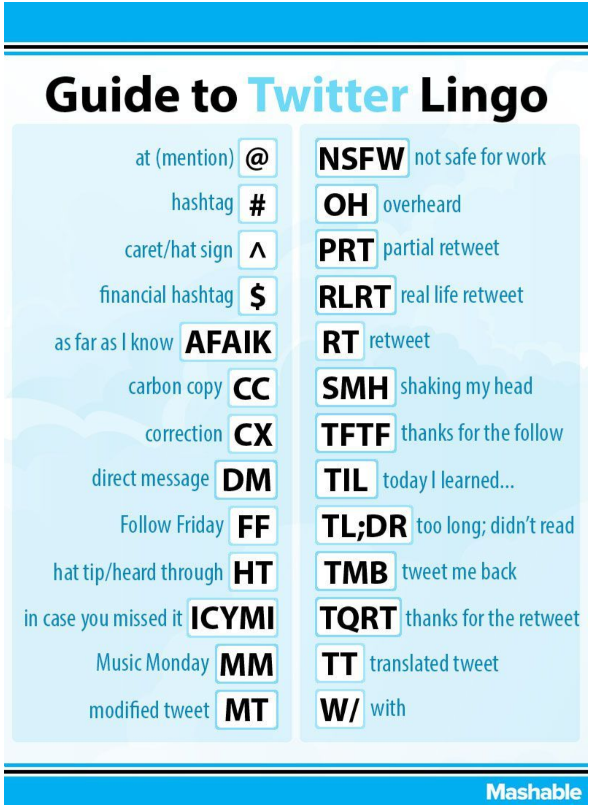
http://sentiment.nrc.ca/lexicons-for-research/

Each word is rated on the scale of happiness

aback	anger	0
aback	anticipation	0
aback	disgust	0
aback	fear	0
aback	joy	0
aback	negative	0
aback	positive	0
aback	sadness	0
aback	surprise	0
aback	trust	0
abacus	anger	0
abacus	anticipation	0
abacus	disgust	0
abacus	fear	0
abacus	joy	0
abacus	negative	0
abacus	positive	0
abacus	sadness	0
abacus	surprise	0
abacus	trust	1
abandon	anger	0
abandon	anticipation	0
abandon	disgust	0
abandon	fear	1
abandon	joy	0
abandon	negative	1
abandon	positive	0
abandon	sadness	1
abandon	surprise	0
abandon	trust	0
abandoned	anger	1
abandoned	anticipation	0
abandoned	disgust	0
abandoned	fear	1

Shortcomings of the dictionary approach

Shortcomings of the dictionary approach



Comments	Internet Slang	Emojis
That's funny →	haha/lol →	😂
Okay	K	👌
I like it	like	👍
Talk to you later	ttyl	👋
Just kidding	jk	😉
I love you	ily	❤️

Shortcomings of the dictionary approach

Sentiment analysis needs interpretations based on communication context.

This tweet is rated as positive by the socialmention algorithm. But do you agree?

How are sentiments detected

The supervised machine approach (will be covered in future classes)

1. Manually label texts by sentiment types and strength
2. Apply machine learning algorithms to the labelled data (also called training data)
3. Algorithms will pick up patterns and rules in human judgement applied to labelling the texts and the algorithms use the patterns to identify sentiment in larger and unlabelled data
4. A series of validation to finetune the algorithm to make it more accurate.

How are sentiments detected

Data Sources

- test.csv 300k x
- train.csv 100.0k x

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	airline	airline_sentiment_gold	name	negativereason_gold	retweet_count	text	tweet_coord	tweet_created	tweet_location	user_timezone	
2	5.70306E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica What @dhep		2015-02-24 11:35:52 -0800		Eastern Time (US & Canada)	
3	5.70301E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plus you've a		2015-02-24 11:15:59 -0800		Pacific Time (US & Canada)	
4	5.70301E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I didn't today		2015-02-24 11	Lets Play	Central Time (US & Canada)	
5	5.70301E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's really agg		2015-02-24 11:15:36 -0800		Pacific Time (US & Canada)	
6	5.70301E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and it's a rea		2015-02-24 11:14:45 -0800		Pacific Time (US & Canada)	
7	5.70301E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@VirginAmerica seriously would pay $30 a flight for seats that didn't have this playing. it's really the only bad thing about flying VA		2015-02-24 11:14:33 -0800		Pacific Time (US & Canada)	
8	5.70301E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes, nearly e		2015-02-24 11	San Francisco C	Pacific Time (US & Canada)	
9	5.703E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Really missed		2015-02-24 11	Los Angeles	Pacific Time (US & Canada)	
	5.703E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Well, I didn't.		2015-02-24 11	San Diego	Pacific Time (US & Canada)	
	5.70295E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it was amazi		2015-02-24 10	Los Angeles	Eastern Time (US & Canada)	
	5.70294E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did you know		2015-02-24 10	1/1 loner squad	Eastern Time (US & Canada)	
13	5.7029E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I <3 pretty		2015-02-24 10	NYC	America/New_York	
	5.7029E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica This is such a		2015-02-24 10	NYC	America/New_York	
	5.70287E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @virginmedi		2015-02-24 10:21:28 -0800		Eastern Time (US & Canada)	
	5.70286E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Thanks!		2015-02-24 10	San Francisco, C	Pacific Time (US & Canada)	
17	5.70282E+17	negative	0.6842	Late Flight	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFO-PDX sch		2015-02-24 10	palo alto, ca	Pacific Time (US & Canada)	
18	5.70278E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So excited fo		2015-02-24 09	west covina	Pacific Time (US & Canada)	
19	5.70277E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I flew from		2015-02-24 09	this place called	Eastern Time (US & Canada)	
20	5.70271E+17	positive	1			Virgin America		thebrandiray		0	I ❤ flying @VirginAmerica. ☺		2015-02-24 09	Somewhere cel	Atlantic Time (Canada)	
21	5.70268E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you know wh		2015-02-24 09	Boston | Walth	Quito	
22	5.70266E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica why are you		2015-02-24 08:55:56 -0800			

The training sample contains human-coded sentiment type of each tweet

https://www.kaggle.com/crowdflower/twitter-airline-sentiment#Tweets.csv

How are sentiments detected

1	text	screen_name	id	polite
2	@NSinclaireMEP Knew that Lib Dems getting into bed with Tories would end like	martinwedge	4.73056E+17	impolite
3	#Skynews.BREAKING-@Nigel_Farage attacked again.C.I.D at the scene.More as v	Comrade58	4.64752E+17	impolite
4	@JaniceUKIP winning? like Charlie Cheen? Do you drink Tiger Blood as well? You	dennisterrey	4.68755E+17	impolite
5	@ostercywriter @Ross_Greer @Mr_JDTraynor @DavidCoburnUKip disgusting m	CrinklyCree	4.67664E+17	impolite
6	@ms_fry @Nigel_Farage the man's a beacon of left-wing, tolerant thought	KingdomOfTheEgo	4.69599E+17	impolite
7	Hahahaha @nickgriffinmep has been hacked, what a bellend	daavlawson	4.62358E+17	impolite
8	@marcuschown Anyone who does vote for him should be shot at dawn, they are	crusader4animal	4.68849E+17	impolite
9	@nickgriffinmep @StanCollymore Fuck off nick	Tom_Stapleton	4.71233E+17	impolite
10	@MikejMcDermott @DavidCoburnUKip Ah. I SEE! You wish to ban anything that	AlanJohnson35	4.69147E+17	impolite
11	@nickcarthew he'll be the one with no hair as he's ripped it out due to Tower Ha	simon4europe	4.70707E+17	impolite
12	@Tim_Aker @UKIP haha what about your mate #NifyNige? You're all thieves. ht	BoxingKangaroo	4.66628E+17	impolite
13	@Nigel_Farage is Britain's Berlusconi ‚Äì someone foreigners look at and say, "Ho	willshome	4.69118E+17	impolite
14	I would quite like to push @Nigel_Farage down those escalater steps http://t.co/	joydxvision	4.70895E+17	impolite
15	RT @helenlewis: Striking result of local elections so far: Ukip on 25% vote share c	Green_DannyB	4.69737E+17	impolite

https://github.com/pablobarbera/eui-text-workshop/blob/master/datasets/EP-elections-tweets.csv

How are sentiments detected

Supervised machine learning approach

- More sensitive to contextual cues of sentiment
- Involves human judgement
- Requires a series of validation

Practice

- Make sure the source code can produce on your machine the same output as you see on the previous page;
- Instead of plotting daily average anger score, let’s create a plot for daily average **anticipation** count.
- Make the code work for your data

Practice script at https://curiositybits.cc/post/r_analytics9/