Finding Happiness

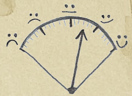
Principles of Complex Systems
CSYS/MATH 300, Spring, 2013 | #SpringPoCS2013

Prof. Peter Dodds
@peterdodds

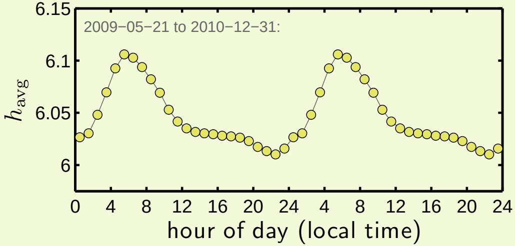
Department of Mathematics & Statistics | Center for Complex Systems | Vermont Advanced Computing Center | University of Vermont

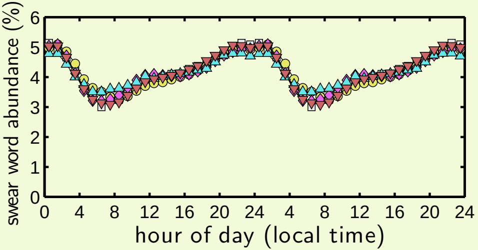
These slides brought to you by:

Outline



The daily unravelling of the human mind:





Happiness:

Measuring Happiness

Data sets

Blogs

Twitter

Geography

Health

Demographics

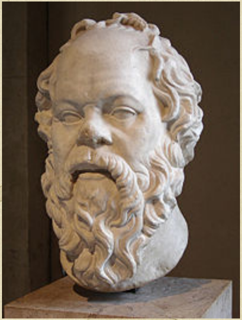
Movement

Networks

Phrases

The End

References



Socrates et al.:
eudaimonia [11]

Bentham:
hedonistic calculus

Jefferson:
...the pursuit of happiness

Early drafts:

that among these are:

Life, ✓

Liberty, ✓ and ?? ~~Money~~?

~~Libations~~ ... ~~Alcohol~~

~~Property~~ ~~Foot-the-ball~~

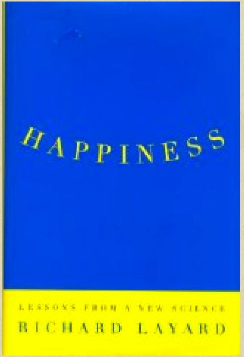
~~Beer~~

Happiness ✓✓✓

Happiness:

Even the odd modern economist is happy:

"Happiness" by Richard Layard [15]



[amazon] (⊞)

What makes us happy?

Layard's summary:

Dominant factors:

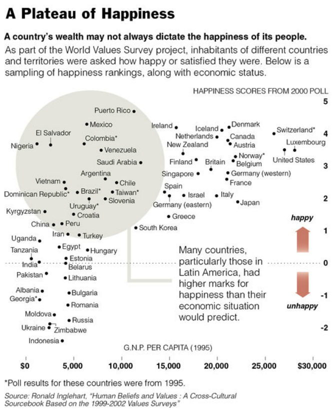
- Family relationships
- Financial situation
- Work
- Community and Friends
- Health
- Personal Values
- Personal Freedom

Unimportant factors:

- Age
- Gender
- Education
- Inherent intelligence
- Looks

Desiring happiness—not just for boffins:

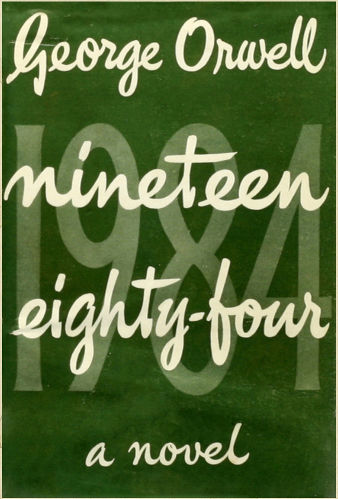
- Average people routinely report being happy is what they want most in life [15, 16, 6]
- And it matters: "Happy people live longer:. . . " Survey by Diener and Chan. [6]



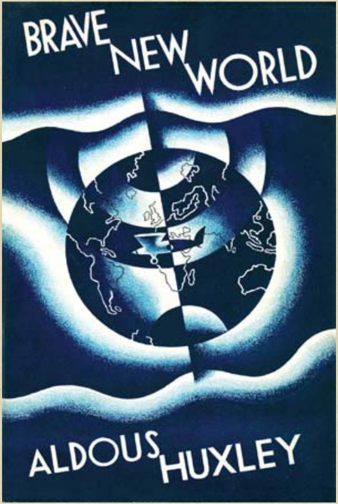
National indices of well-being:

- Bhutan
- France
- Australia

An easy knock:



Science = Orwell



Policy = Brave New World

Measuring Happiness
Data sets
Blogs
Twitter
Geography
Health
Demographics
Movement
Networks
Phrases
The End
References

Emotional content

So how does one measure

1. happiness?
2. levels of other emotional states?

Just ask people how happy they are.

- Experience sampling [3, 5, 4] (Csikszentmihalyi et al.)
- Day reconstruction [12] (Kahneman et al.)

But self-reporting has some drawbacks:

- relies on memory and self-perception
- induces misreporting [17]
- costly

Happiness, attention, and doing:

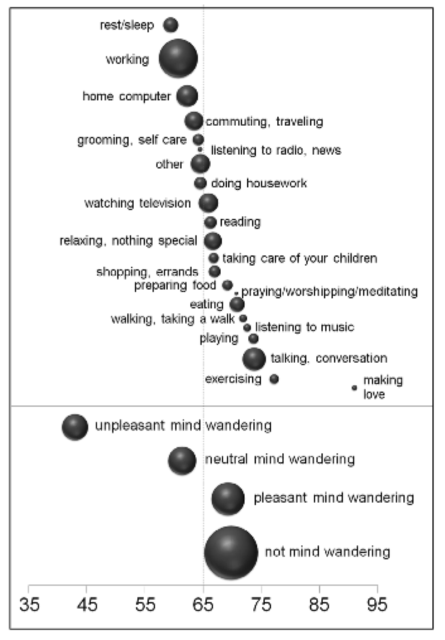
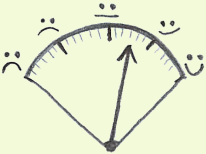


Fig. 1. Mean happiness reported during each activity (**top**) and while mind wandering to unpleasant topics, neutral topics, pleasant topics or not mind wandering (**bottom**). Dashed line indicates mean of happiness across all samples. Bubble area indicates the frequency of occurrence. The largest bubble ("not mind wandering") corresponds to 53.1% of the samples, and the smallest bubble ("praying/worshipping/meditating") corresponds to 0.1% of the samples.

Killingsworth and Gilbert, Science, 2010 [13]

We'd like to build an 'hedonometer':



- An instrument to 'remotely-sense' emotional states and levels, in real time or post hoc.

Ideally:

- Transparent
- Fast
- Based on written expression
- Uses human evaluation
- Non-reactive
- Complementary to self-reported measures
- Improvable

Measuring Emotional Content

- Idea: Build on measures of the emotional content of individual words.
- Osgood et al. (1957) [20] identified a basis of three psychological variables as semantic differentials:
  - Valence: bad $\leftrightarrow$ good
  - Dominance: weak $\leftrightarrow$ strong
  - Arousal: passive $\leftrightarrow$ active

ANEW study

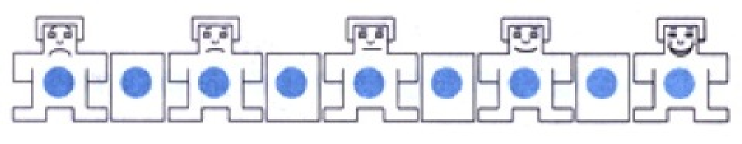
- ANEW = “Affective Norms for English Words”

- Study: participants shown lists of isolated words
- Asked to grade each word’s valence, arousal, and dominance level
- Integer scale of 1–9

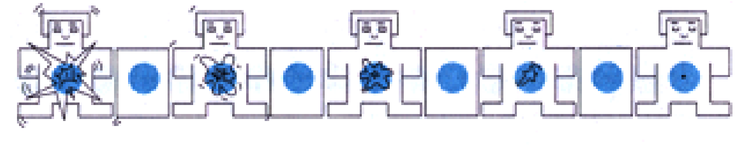
- N =1034 words—previously identified as bearing emotional weight
- Participants = College students (*cough*)
- Results published by Bradley and Lang (1999) [2]

1999 ANEW study—three 1–9 scales: [2]

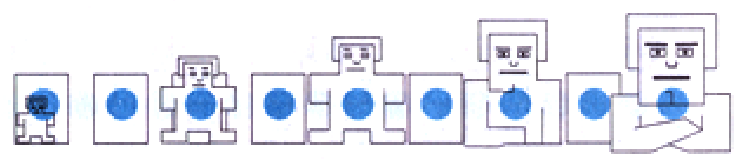
valence:



arousal:



dominance:



ANEW study:

Valence = Happiness:

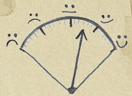
- Valence scale presented to participants as a 'happy-unhappy scale.'
- Participants were further told:

  "At one extreme of this scale, you are happy, pleased, satisfied, contented, hopeful. . . .

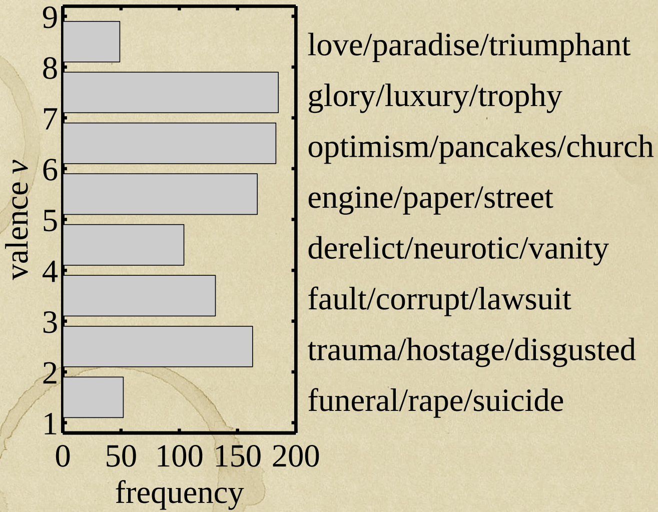
  The other end of the scale is when you feel completely unhappy, annoyed, unsatisfied, melancholic, despaired, or bored."

Top and Bottom 5 words by valence

1	triumphant (8.82)	rape (1.25)
2	paradise (8.72)	suicide (1.25)
3	love (8.72)	funeral (1.39)
4	loved (8.64)	cancer (1.50)
5	miracle (8.60)	rejected (1.50)



ANEW study words—examples



ANEW = "Affective Norms for English Words" [2]

Measuring the happiness of a text:

Lyrics for Michael Jackson's Billie Jean

"She was more like a beauty queen
from a movie scene.
⋮
And mother always told me,
be careful who you love.
And be careful of what you do
'cause the lie becomes the truth.

Billie Jean is not my lover,
She's just a girl who claims
that I am the one.
⋮

ANEW words	v_k	f_k
k=1. love	8.72	1
2. mother	8.39	1
3. baby	8.22	3
4. beauty	7.82	1
5. truth	7.80	1
6. people	7.33	2
7. strong	7.11	1
8. young	6.89	2
9. girl	6.87	4
10. movie	6.86	1
11. perfume	6.76	1
12. queen	6.44	1
13. name	5.55	1
14. lie	2.79	1

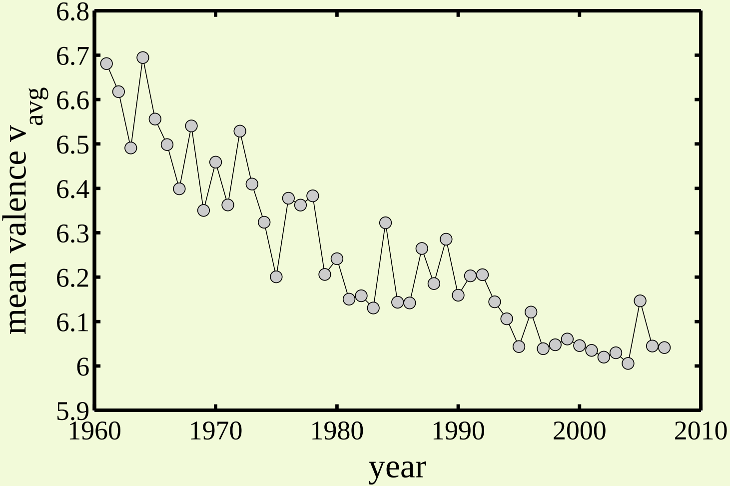
$$v_{\text{text}} = \frac{\sum_k v_k f_k}{\sum_k f_k}$$

$$v_{\text{Billie Jean}} = 7.1$$

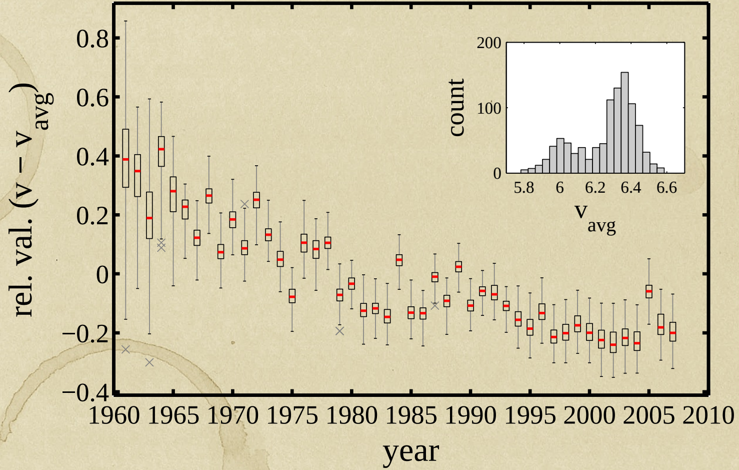
$$v_{\text{Thriller}} = 6.3$$

$$v_{\text{Michael Jackson}} = 6.4$$

Song Lyrics—average happiness

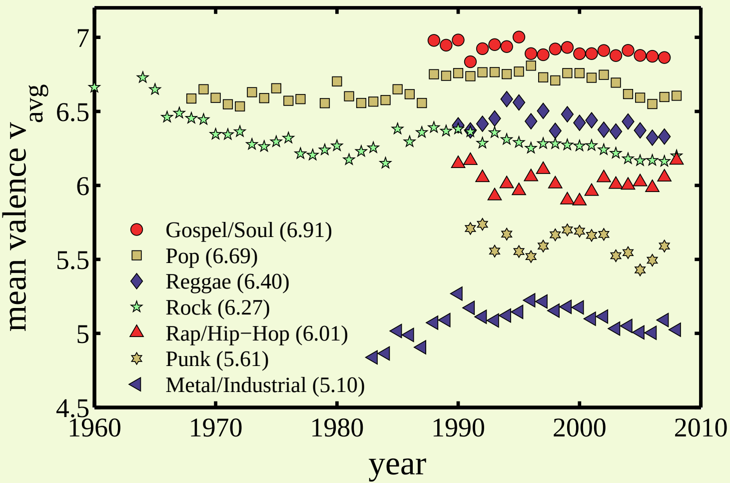


Song Lyrics—measurement robustness



100 random subsets of 750 ANEW words

Song Lyrics—average happiness of genres:



Valence shift details:

Given two texts *a* and *b*:

- Measure difference in average valence: $v_{avg}^{(b)} - v_{avg}^{(a)}$
- Break difference down by contributions from individual words:

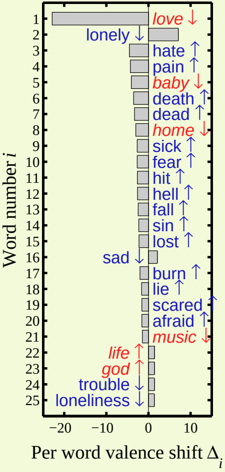
$$\Delta_i = 100 \times [p_{i,b} - p_{i,a}] \frac{[v_i - v_{\text{avg}}^{(a)}]}{[v_{\text{avg}}^{(b)} - v_{\text{avg}}^{(a)}]}$$

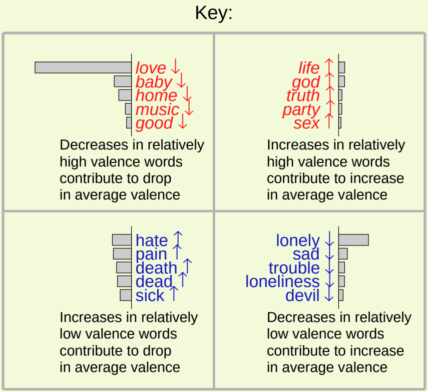
$$\sum_i \Delta_i = v_{avg}^{(b)} - v_{avg}^{(a)}$$

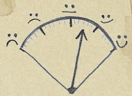
- Rank words by $|\Delta_i|$

Happiness Word Shift Graph:

Per word drop in valence of lyrics from 1980−2007 relative to valence of lyrics from 1960







Top 16 of $\simeq$ 20,000 artists:

Rank	Artist	Valence
1	All-4-One	7.15
2	Luther Vandross	7.12
3	S Club 7	7.05
4	K Ci & JoJo	7.04
5	Perry Como	7.04
6	Diana Ross & The Supremes	7.03
7	Buddy Holly	7.02
8	Faith Evans	7.01
9	The Beach Boys	7.01
10	Jon B	6.98
11	Dru Hill	6.96
12	Earth Wind & Fire	6.95
13	Ashanti	6.95
14	Otis Redding	6.93
15	Faith Hill	6.93
16	NSync	6.93

(criteria: $\geq$ 50 songs and $\geq$ 1000 ANEW words)

Bottom 16 of $\simeq$ 20,000 artists:

Rank	Artist	Valence
1	Slayer	4.80
2	Misfits	4.88
3	Staind	4.93
4	Slipknot	4.98
5	Darkthrone	4.98
6	Death	5.02
7	Black Label Society	5.05
8	Pig	5.08
9	Voivod	5.14
10	Fear Factory	5.15
11	Iced Earth	5.16
12	Simple Plan	5.16
13	Machine Head	5.17
14	Metallica	5.19
15	Dimmu Borgir	5.20
16	Mudvayne	5.21

(criteria: $\geq$ 50 songs and $\geq$ 1000 ANEW words)

Data sets:

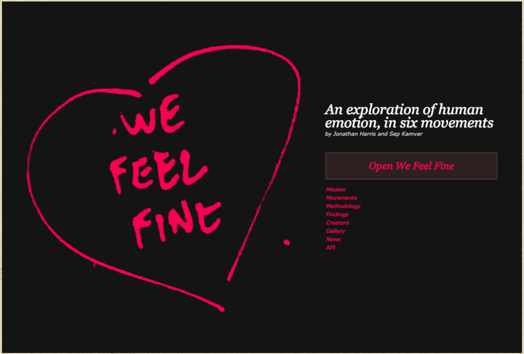
Texts:

- Song lyrics and titles (1960–2008)
- State of the Union (SOTU) Addresses (1790–2008)
- Twitter, 2008—
- Blogs (wefeelfine.org)
- New York Times (20 years)
- Gutenberg.org
- Google Books: http://ngrams.googlelabs.com/ (⊞)
- BBC transcripts

⋮

Data sets:

- Blog phrases containing “I feel...”, “I am feeling”, etc., taken from wefeelfine.org (⊞) (API, 2005–2010)

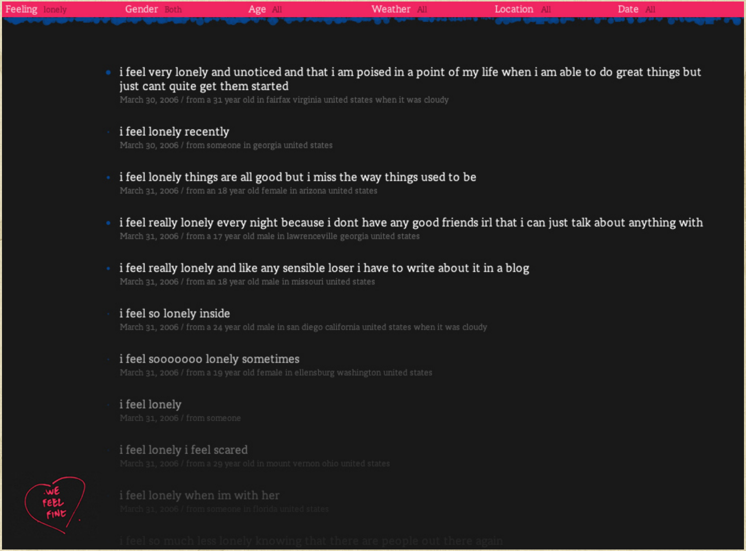



- Created by Jonathan Harris & Sep Kamvar

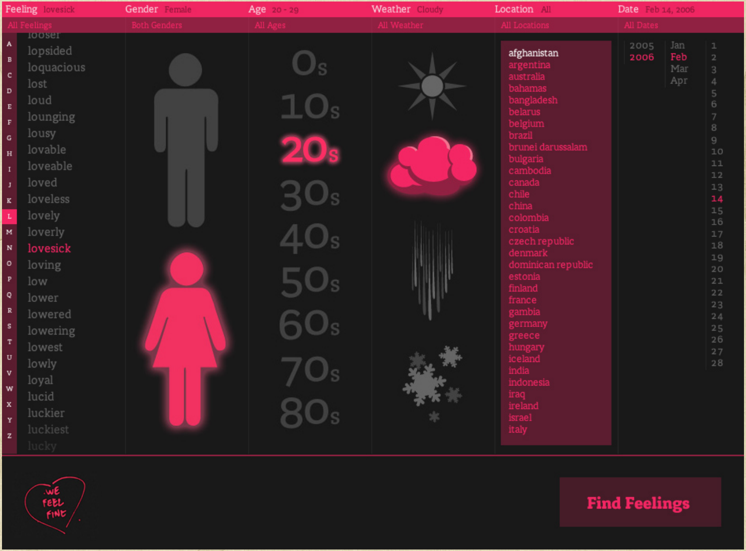
Jonathan Harris, wefeelfine.org

(Loading Movie)

wefeelfine.org:



wefeelfine.org:



More data sets:

5.
6. New York Times (20 years)
7. Gutenberg.org
8. Google Books: http://ngrams.googlelabs.com/ (⊞)
9. …

Data sets:

Counts	Song lyrics	Song titles
All words	58,610,849	60,867,223
Individuals	$\sim$ 20,000	$\sim$ 632,000

Counts	blogs	SOTU
All words	155,667,394	1,796,763
Individuals	$\sim$ 2,335,000	43

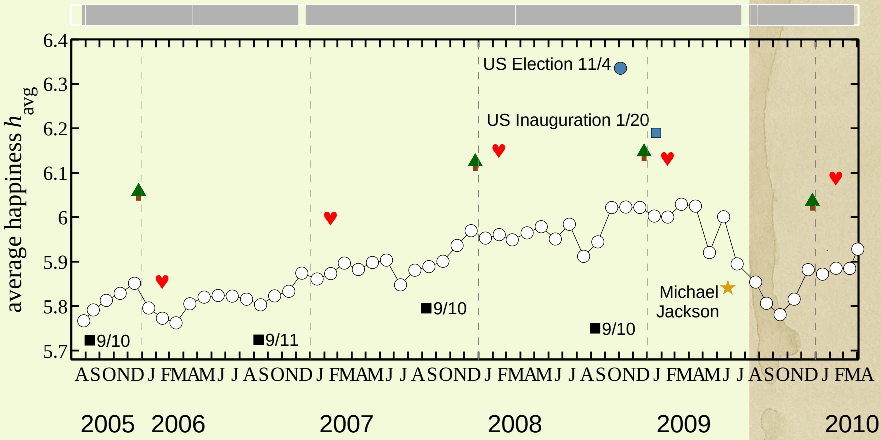
Counts	Twitter
All words	$\sim$ 100 billion
Tweets	$\sim$ 10 billion
Individuals	$\sim$ 100 million

Most frequent ANEW words:

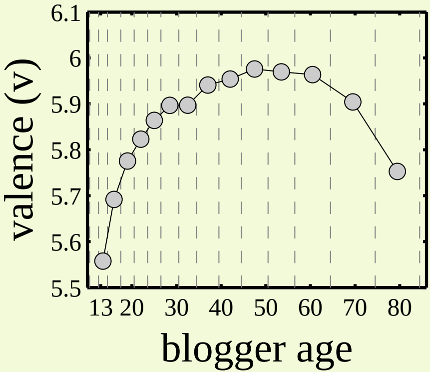
Rank	Song lyrics	Song titles
1	love (7.37%)	love (7.39%)
2	time (4.18%)	time (4.19%)
3	baby (2.75%)	baby (2.75%)
4	life (2.59%)	life (2.60%)
5	heart (2.14%)	heart (2.15%)

Rank	blogs	SOTU	twitter
1	good (4.89%)	people (5.49%)	good (4.50%)
2	time (4.72%)	time (4.09%)	love (4.45%)
3	people (3.94%)	present (3.45%)	time (3.30%)
4	love (3.31%)	world (3.10%)	people (2.06%)
5	life (3.13%)	war (2.98%)	home (1.71%)

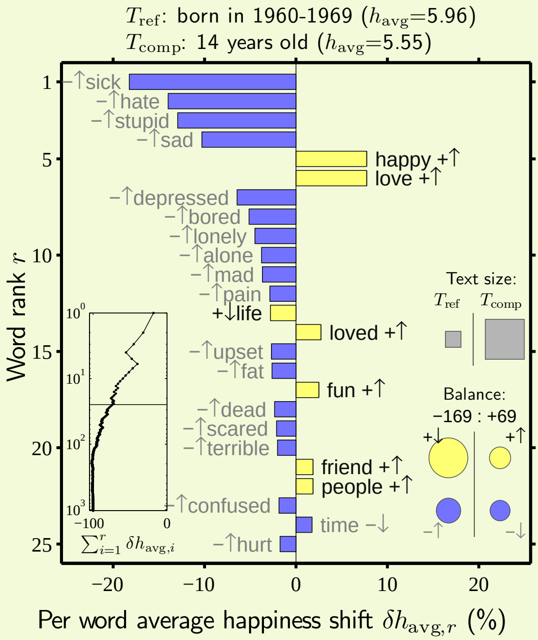
Blogs—Overall trend



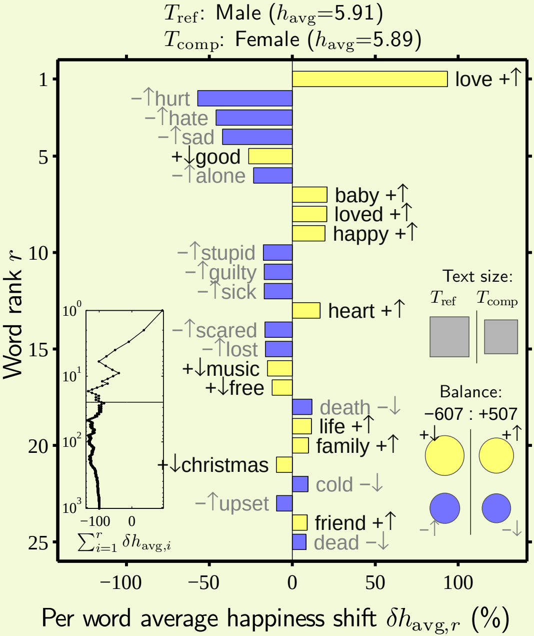
Blogs—Age:



- Average happiness as a function of the age bloggers report they will turn in the year of their posting.



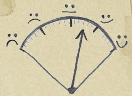
Measuring Happiness
Data sets
Blogs
Twitter
Geography
Health
Demographics
Movement
Networks
Phrases
The End
References



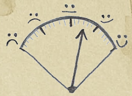
Measuring Happiness
Data sets
Blogs
Twitter
Geography
Health
Demographics
Movement
Networks
Phrases
The End
References

labMT 1.0: language assessment by Mechanical Turk

- Twitter, Google Books, Music Lyrics, and the New York Times.
- 5000 most frequency used words for each corpus.
- 10,222 words, 50 evaluations each.



valence rank	word	valence	std dev	twitter rank	g-books rank	nyt rank	lyrics rank
1	laughter	8.50	0.93	3600	–	–	1728
2	happiness	8.44	0.97	1853	2458	–	1230
3	love	8.42	1.11	25	317	328	23
4	happy	8.30	0.99	65	1372	1313	375
5	laughed	8.26	1.16	3334	3542	–	2332
6	laugh	8.22	1.37	1002	3998	4488	647
7	laughing	8.20	1.11	1579	–	–	1122
8	excellent	8.18	1.10	1496	1756	3155	–
9	laughs	8.18	1.16	3554	–	–	2856
10	joy	8.16	1.06	988	2336	2723	809
11	successful	8.16	1.08	2176	1198	1565	–
12	win	8.12	1.08	154	3031	776	694
13	rainbow	8.10	0.99	2726	–	–	1723
14	smile	8.10	1.02	925	2666	2898	349
15	won	8.10	1.22	810	1167	439	1493
16	pleasure	8.08	0.97	1497	1526	4253	1398
17	smiled	8.08	1.07	–	3537	–	2248
18	rainbows	8.06	1.36	–	–	–	4216
19	winning	8.04	1.05	1876	–	1426	3646
20	celebration	8.02	1.53	3306	–	2762	4070
21	enjoyed	8.02	1.53	1530	2908	3502	–
22	healthy	8.02	1.06	1393	3200	3292	4619
23	music	8.02	1.12	132	875	167	374
24	celebrating	8.00	1.14	2550	–	–	–
25	congratulations	8.00	1.63	2246	–	–	–
26	weekend	8.00	1.29	317	–	833	2256
27	celebrate	7.98	1.15	1606	–	3574	2108
28	comedy	7.98	1.15	1444	–	2566	–
29	jokes	7.98	0.98	2812	–	–	3808
30	rich	7.98	1.32	1625	1221	1469	890
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

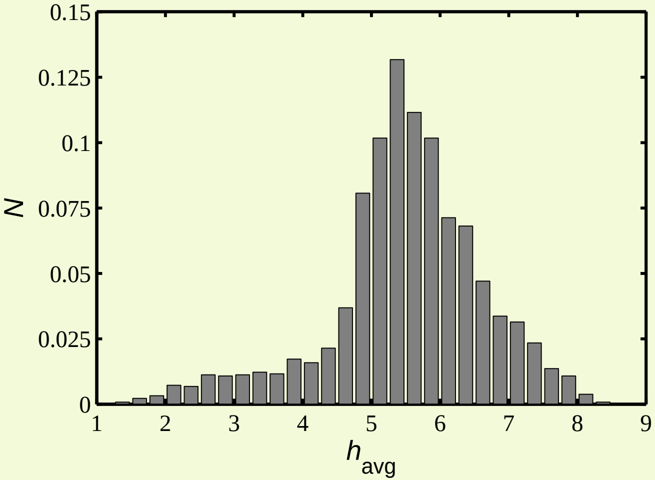


The UNIVERSITY of VERMONT

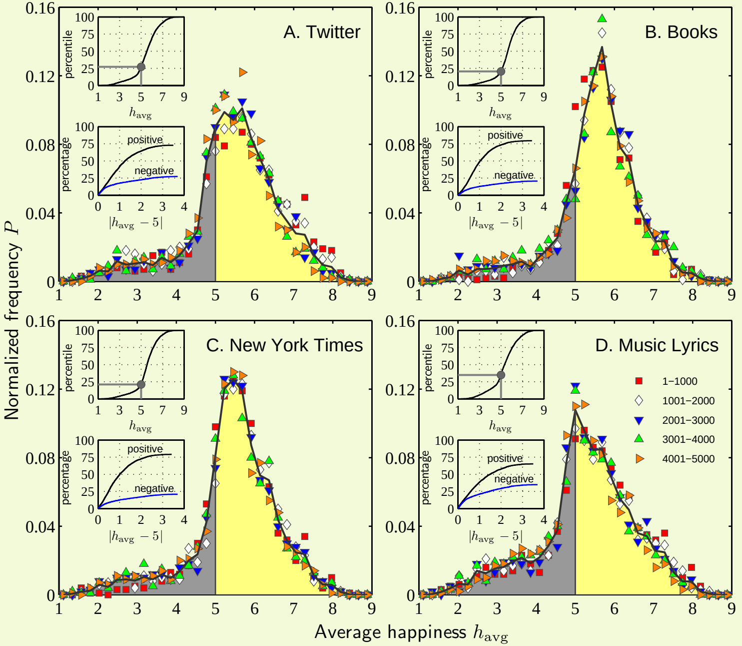
valence rank	word	valence	std dev	twitter rank	g-books rank	nyt rank	lyrics rank
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10193	violence	1.86	1.05	4299	1724	1238	2016
10194	cruel	1.84	1.15	2963	–	–	1447
10195	cry	1.84	1.28	1028	3075	–	226
10196	failed	1.84	1.00	2645	1618	1276	2920
10197	sickness	1.84	1.18	4735	–	–	3782
10198	abused	1.83	1.31	–	–	–	4589
10199	tortured	1.82	1.42	–	–	–	4693
10200	fatal	1.80	1.53	–	4089	–	3724
10201	killings	1.80	1.54	–	–	4914	–
10202	murdered	1.80	1.63	–	–	–	4796
10203	war	1.80	1.41	468	175	291	462
10204	kills	1.78	1.23	2459	–	–	2857
10205	jail	1.76	1.02	1642	–	2573	1619
10206	terror	1.76	1.00	4625	4117	4048	2370
10207	die	1.74	1.19	418	730	2605	143
10208	killing	1.70	1.36	1507	4428	1672	998
10209	arrested	1.64	1.01	2435	4474	1435	–
10210	deaths	1.64	1.14	–	–	2974	–
10211	raped	1.64	1.43	–	–	–	4528
10212	torture	1.58	1.05	3175	–	–	3126
10213	died	1.56	1.20	1223	866	208	826
10214	kill	1.56	1.05	798	2727	2572	430
10215	killed	1.56	1.23	1137	1603	814	1273
10216	cancer	1.54	1.07	946	1884	796	3802
10217	death	1.54	1.28	509	307	373	433
10218	murder	1.48	1.01	2762	3110	1541	1059
10219	terrorism	1.48	0.91	–	–	3192	–
10220	rape	1.44	0.79	3133	–	4115	2977
10221	suicide	1.30	0.84	2124	4707	3319	2107
10222	terrorist	1.30	0.91	3576	–	3026	–

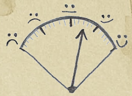
std dev rank	word	valence	std dev	twitter rank	g-books rank	nyt rank	lyrics rank
1	f↯@king	4.64	2.93	448	–	–	620
2	f✩✩kin	3.86	2.74	1077	–	–	688
3	f✩✩ked	3.56	2.71	1840	–	–	904
4	pussy	4.80	2.66	2019	–	–	949
5	whiskey	5.72	2.64	–	–	–	2208
6	slut	3.57	2.63	–	–	–	4071
7	cigarettes	3.31	2.60	–	–	–	3279
8	f✩✩k	4.14	2.58	322	–	–	185
9	mortality	4.38	2.55	–	3960	–	–
10	cigarette	3.09	2.52	–	–	–	2678
11	motherf✩✩kers	2.51	2.47	–	–	–	1466
12	churches	5.70	2.46	–	2281	–	–
13	motherf✩✩king	2.64	2.46	–	–	–	2910
14	capitalism	5.16	2.45	–	4648	–	–
15	porn	4.18	2.43	1801	–	–	–
16	summer	6.40	2.39	896	1226	721	590
17	beer	5.92	2.39	839	4924	3960	1413
18	execution	3.10	2.39	–	2975	–	–
19	wines	6.28	2.37	–	–	3316	–
20	zombies	4.00	2.37	4708	–	–	–
21	aids	4.28	2.35	2983	3996	1197	–
22	capitalist	4.84	2.34	–	4694	–	–
23	revenge	3.71	2.34	–	–	–	2766
24	mcdonalds	5.98	2.33	3831	–	–	–
25	beatles	6.44	2.33	3797	–	–	–
26	islam	4.68	2.33	–	4514	–	–
27	pay	5.30	2.32	627	769	460	499
28	alcohol	5.20	2.32	2787	2617	3752	3600
29	muthaf✩✩kin	3.00	2.31	–	–	–	4107
30	christ	6.16	2.31	2509	909	4238	1526
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

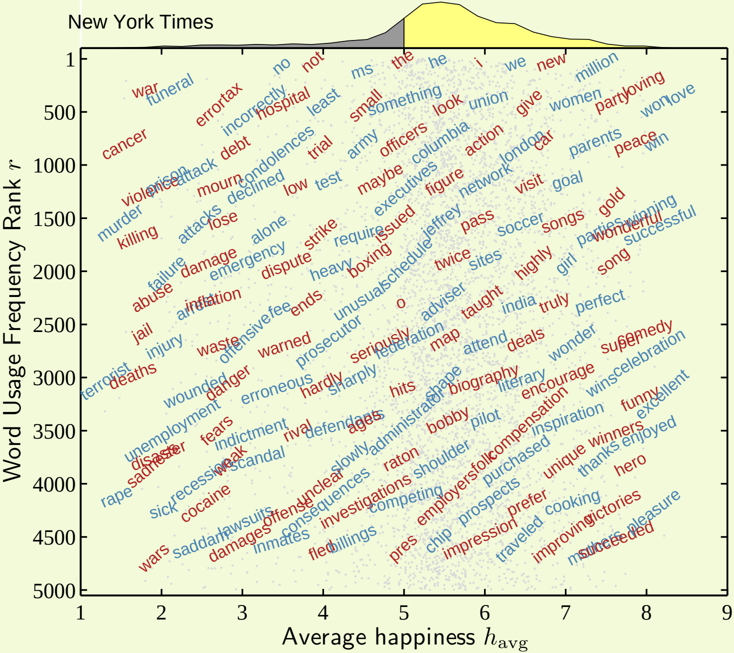
English's scale-invariant, positive bias:[14]

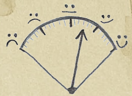


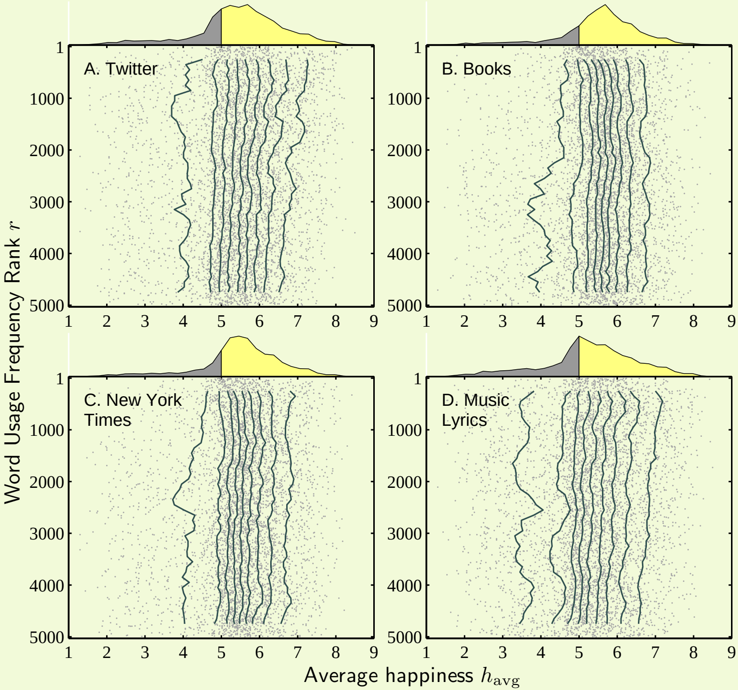
- Social organism story manifested in language.


A. Twitter
B. Books
C. New York Times
D. Music Lyrics
Normalized frequency P
Average happiness $h_{\rm avg}$
percentile
$h_{\rm avg}$
percentage
positive
negative
$|h_{\rm avg} - 5|$
1–1000
1001–2000
2001–3000
3001–4000
4001–5000

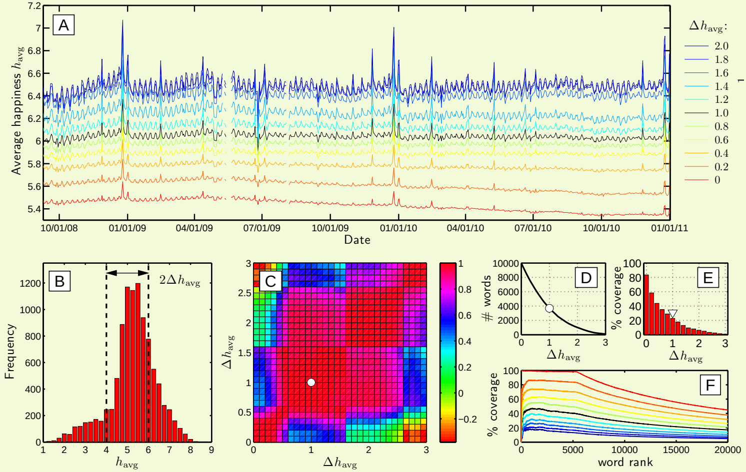


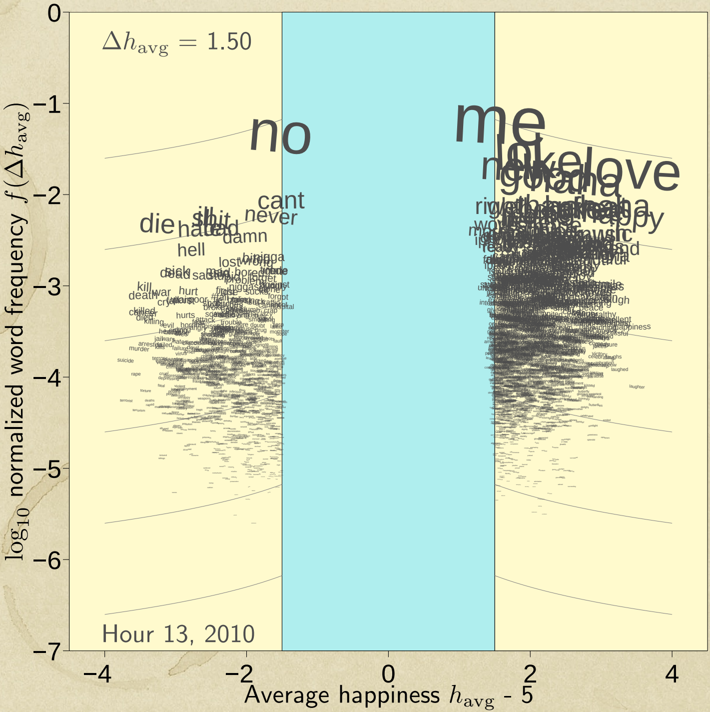

New York Times
Word Usage Frequency Rank r
Average happiness h_{avg}
1
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
1 2 3 4 5 6 7 8 9
war funeral no not ms the he i we new million
errortax incorrectly hospital least small something look union give women party loving won love
cancer debt condolences trial army officers columbia action london car parents peace win
violence prison attack mourn declined low test maybe executives figure network visit goal
murder killing attacks lose alone strike require issued jeffrey pass soccer songs gold parties winning wonderful successful
failure damage emergency dispute heavy boxing schedule twice sites highly girl song
abuse inflation affiliation offensive fee ends unusual o adviser taught india truly perfect
jail injury waste warned prosecutor seriously federation map attend deals wonder super comedy
terrorist deaths wounded danger erroneous hardly sharply hits shape biography literary encourage wins celebration
unemployment fears indictment rival defendants ages administrator bobby pilot compensation inspiration funny excellent
disaster sadness rape weak scandal recession slowly raton shoulder purchased unique winners thanks enjoyed hero
sick cocaine unclear consequences investigations competing employers folk prospects prefer cooking victories pleasure
wars saddam lawsuits damages offense inmates fled billings pres chip impression traveled improving members succeeded



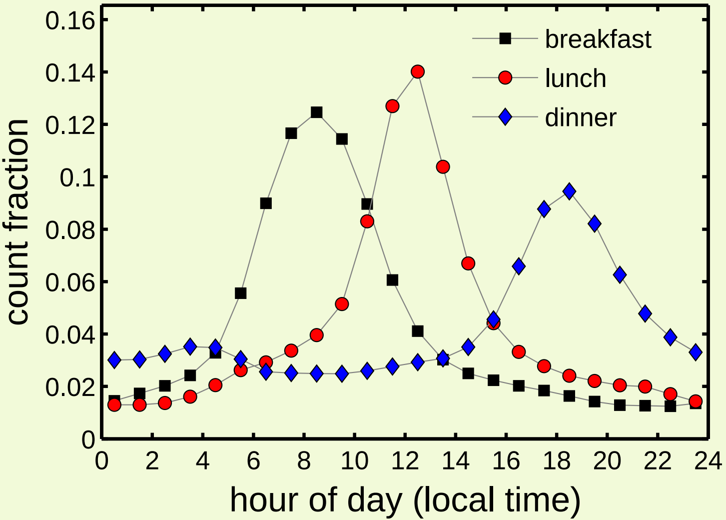

A. Twitter
B. Books
C. New York Times
D. Music Lyrics
Word Usage Frequency Rank r
Average happiness h_{avg}
1
1000
2000
3000
4000
5000
1 2 3 4 5 6 7 8 9

The very surprising tunable hedonometer:



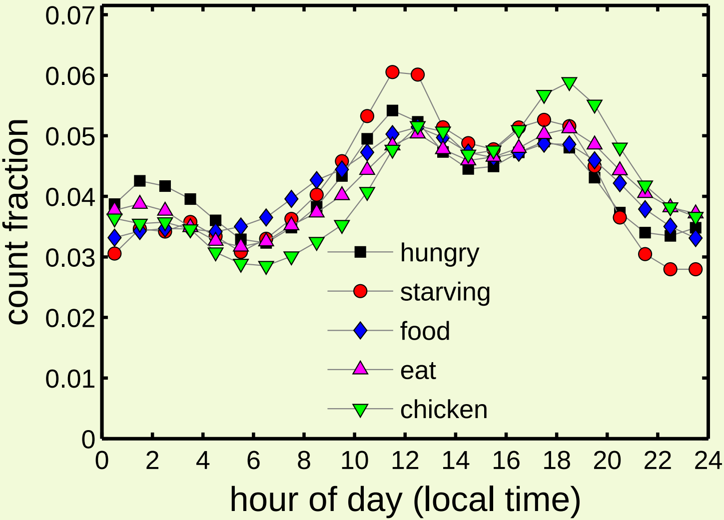

$\Delta h_{\mathrm{avg}} = 1.50$
Hour 13, 2010
no
me
love
cant
never
die
damn
hell
laughter
$\log_{10}$ normalized word frequency $f(\Delta h_{\mathrm{avg}})$
Average happiness h_{avg} - 5
0
−1
−2
−3
−4
−5
−6
−7
−4
−2
0
2
4

Twitter—living in the now:



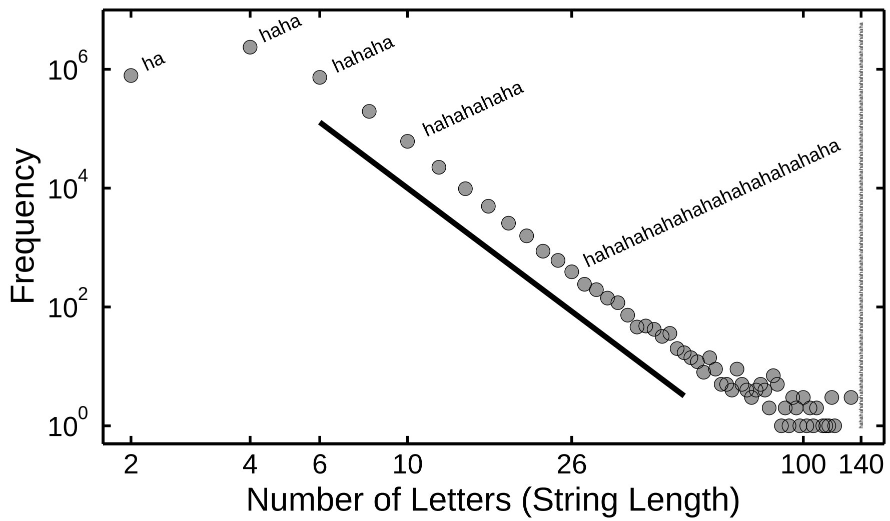
- Quantifying the quotidian.

Twitter—living in the now:

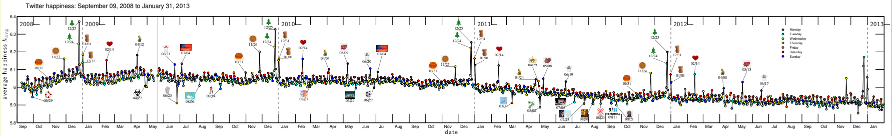


- Makes the unexpected believable...

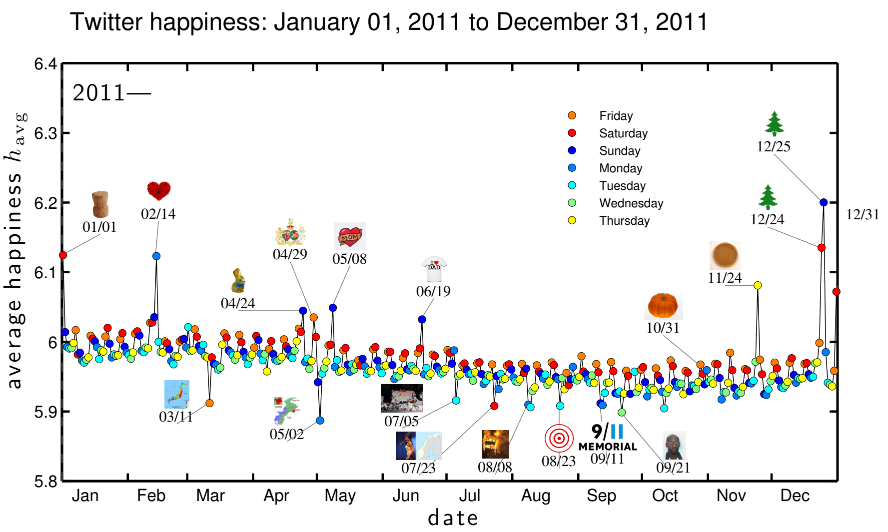
The happiest distribution:



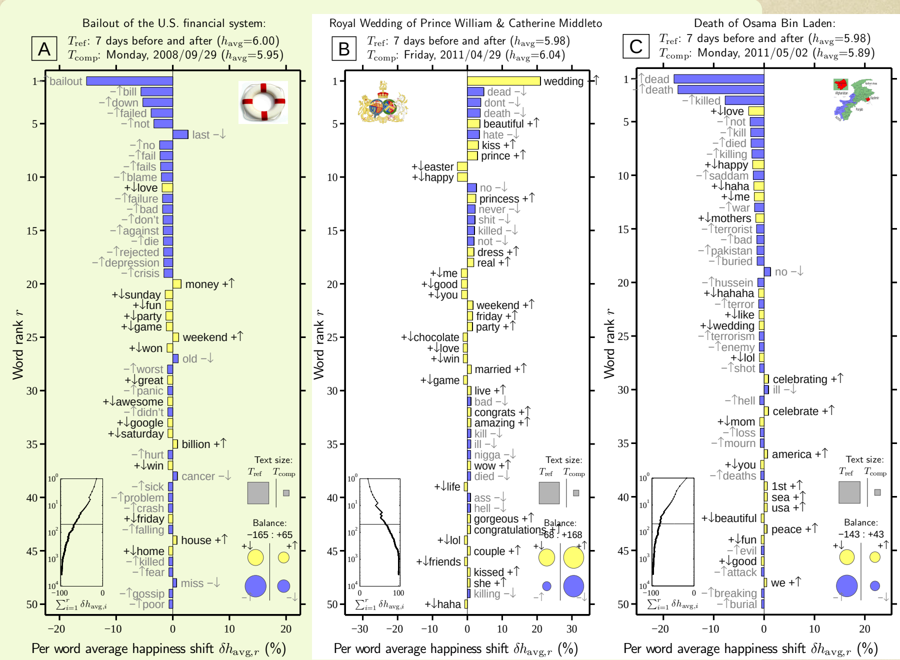
Twitter—overall time series:



- Global happiness spikes due to predictable rituals.
- Global sadness spikes due to unpredictable, exogeneous shocks.

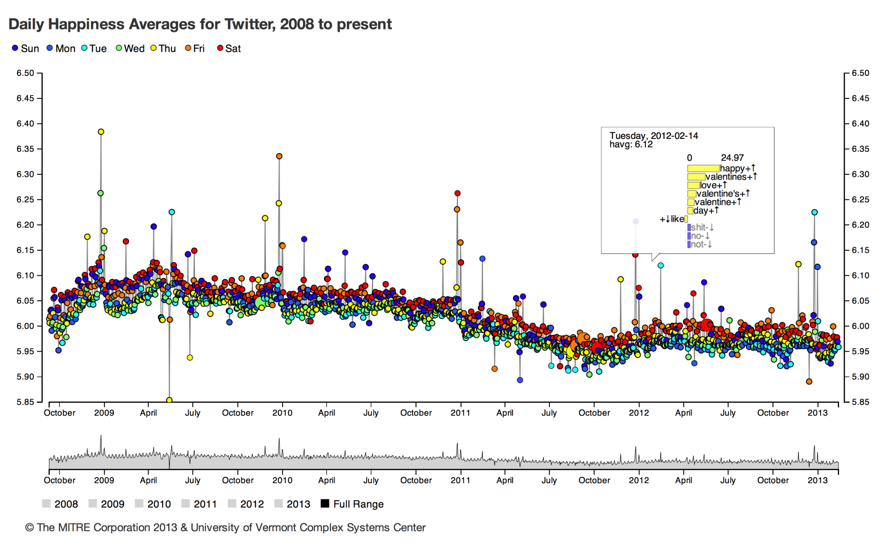


Twitter happiness: January 01, 2011 to December 31, 2011
2011—
average happiness h_{avg}
6.4
6.3
6.2
6.1
6
5.9
5.8
Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
01/01
02/14
03/11
04/24
04/29
05/02
05/08
06/19
07/05
07/23
08/08
08/23
9/11
MEMORIAL
09/11
09/21
10/31
11/24
12/24
12/25
12/31
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
date



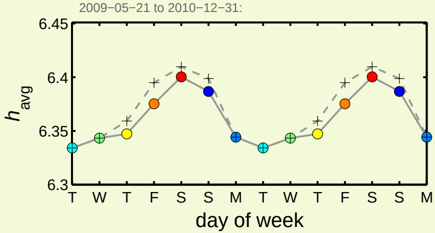
A
Bailout of the U.S. financial system:
T_{ref}: 7 days before and after (h_{avg}=6.00)
T_{comp}: Monday, 2008/09/29 (h_{avg}=5.95)
−↑bailout
−↑bill
−↑down
−↑failed
−↑not
last −↓
−↑no
−↑fail
−↑fails
−↑blame
+↓love
−↑failure
−↑bad
−↑don't
−↑against
−↑die
−↑rejected
−↑depression
−↑crisis
money +↑
+↓sunday
+↓fun
+↓party
+↓game
weekend +↑
+↓won
old −↓
−↑worst
+↓great
−↑panic
+↓awesome
−↑didn't
+↓google
+↓saturday
billion +↑
−↑hurt
+↓win
cancer −↓
−↑sick
−↑problem
−↑crash
+↓friday
−↑falling
house +↑
+↓home
−↑killed
−↑fear
miss −↓
−↑gossip
−↑poor
Text size:
T_{ref} T_{comp}
Balance:
−165 : +65
B
Royal Wedding of Prince William & Catherine Middleto
T_{ref}: 7 days before and after (h_{avg}=5.98)
T_{comp}: Friday, 2011/04/29 (h_{avg}=6.04)
wedding +↑
dead −↓
dont −↓
death −↓
beautiful +↑
hate −↓
kiss +↑
prince +↑
+↓easter
+↓happy
no −↓
princess +↑
never −↓
shit −↓
killed −↓
not −↓
dress +↑
real +↑
+↓me
+↓good
+↓you
weekend +↑
friday +↑
party +↑
+↓chocolate
+↓love
+↓win
married +↑
+↓game
live +↑
bad −↓
congrats +↑
amazing +↑
kill −↓
ill −↓
nigga −↓
wow +↑
died −↓
+↓life
ass −↓
hell −↓
gorgeous +↑
congratulations +↑
+↓lol
couple +↑
+↓friends
kissed +↑
she +↑
killing −↓
+↓haha
Text size:
T_{ref} T_{comp}
Balance:
−68 : +168
C
Death of Osama Bin Laden:
T_{ref}: 7 days before and after (h_{avg}=5.98)
T_{comp}: Monday, 2011/05/02 (h_{avg}=5.89)
−↑dead
−↑death
−↑killed
+↓love
−↑not
−↑kill
−↑died
−↑killing
+↓happy
−↑saddam
+↓haha
+↓me
−↑war
+↓mothers
−↑terrorist
−↑bad
−↑pakistan
−↑buried
no −↓
−↑hussein
+↓hahaha
−↑terror
+↓like
+↓wedding
−↑terrorism
−↑enemy
+↓lol
−↑shot
celebrating +↑
ill −↓
−↑hell
celebrate +↑
+↓mom
−↑loss
−↑mourn
america +↑
+↓you
−↑deaths
1st +↑
sea +↑
usa +↑
+↓beautiful
peace +↑
+↓fun
−↑evil
+↓good
−↑attack
we +↑
−↑breaking
−↑burial
Text size:
T_{ref} T_{comp}
Balance:
−143 : +43
Word rank r
Per word average happiness shift $\delta h_{\text{avg},r}$ (%)
$\sum_{i=1}^{r} \delta h_{\text{avg},i}$

hedonometer.org (⊞) (launching Tuesday, April 30, 2013)

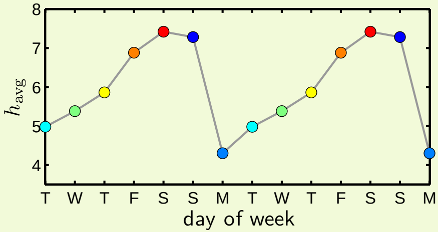


Twitter—weekly time series:

What people say:



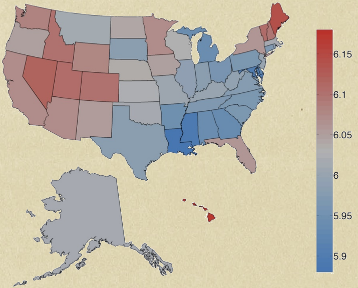
What people think:



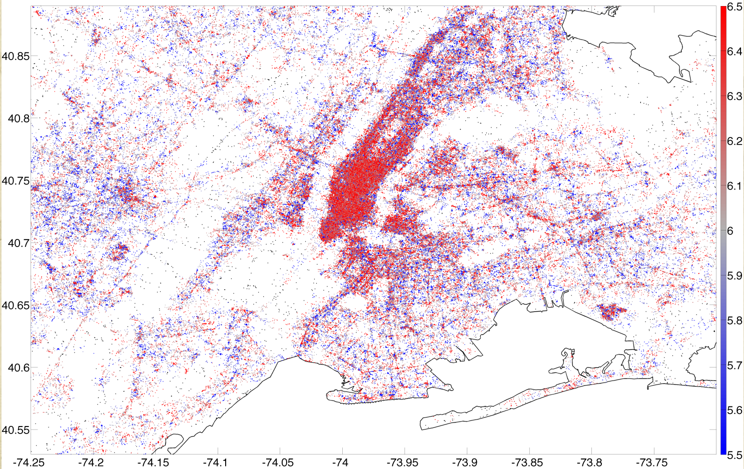
- Inflation: NYT piece (⊞) on the blueness of Tuesdays.

Word	$h_{\mathrm{avg}}^{(\mathrm{amb})}$	Total Tweets	$h_{\mathrm{avg}}^{(\mathrm{norm})}$	Word	$h_{\mathrm{avg}}^{(\mathrm{amb})}$	Total Tweets	$h_{\mathrm{avg}}^{(\mathrm{norm})}$
1. happy	+0.430	1.65e+07 (13)	+1.104 (1)	51. snow	-0.051	2.60e+06 (49)	+0.083 (39)
2. Christmas	+0.404	4.89e+06 (35)	+0.953 (3)	52. Jon Stewart	-0.052	5.21e+04 (97)	-0.024 (48)
3. vegan	+0.315	1.84e+05 (90)	-0.015 (46)	53. school	-0.056	9.26e+06 (24)	+0.050 (42)
4. :)	+0.274	1.04e+07 (20)	+0.630 (12)	54. Lehman Brothers	-0.078	8.50e+03 (100)	-0.721 (79)
5. family	+0.251	5.01e+06 (32)	+0.716 (7)	55. them	-0.090	1.54e+07 (15)	-0.280 (60)
6. :-)	+0.228	1.67e+06 (60)	+0.560 (16)	56. right	-0.090	1.92e+07 (10)	+0.126 (35)
7. our	+0.207	1.41e+07 (16)	+0.159 (33)	57. woman	-0.115	2.54e+06 (51)	+0.202 (30)
8. win	+0.204	7.98e+06 (26)	+0.924 (4)	58. left	-0.118	4.89e+06 (34)	-0.383 (63)
9. vacation	+0.200	9.35e+05 (67)	+0.817 (5)	59. me	-0.119	1.44e+08 (4)	+0.160 (32)
10. party	+0.170	6.44e+06 (29)	+0.679 (9)	60. election	-0.127	5.60e+05 (75)	-0.306 (61)
11. love	+0.164	4.67e+07 (6)	+0.977 (2)	61. Sarah Palin	-0.128	2.26e+05 (87)	-0.681 (76)
12. friends	+0.155	7.67e+06 (27)	+0.685 (8)	62. no	-0.132	9.51e+07 (5)	-1.415 (90)
13. hope	+0.149	1.18e+07 (18)	+0.515 (19)	63. rain	-0.134	3.23e+06 (41)	+0.050 (44)
14. coffee	+0.147	2.80e+06 (46)	+0.518 (18)	64. climate	-0.135	3.64e+05 (80)	-0.160 (51)
15. cash	+0.146	1.28e+06 (63)	+0.601 (14)	65. gay	-0.152	2.73e+06 (47)	-0.552 (72)
16. sun	+0.144	2.39e+06 (52)	+0.737 (6)	66. lose	-0.157	2.06e+06 (55)	-1.181 (86)
17. income	+0.137	5.10e+05 (76)	+0.621 (13)	67. they	-0.159	2.74e+07 (8)	-0.208 (58)
18. summer	+0.135	3.00e+06 (43)	+0.221 (29)	68. oil	-0.162	1.38e+06 (62)	-0.411 (65)
19. church	+0.131	1.81e+06 (58)	-0.016 (47)	69. cold	-0.162	3.67e+06 (36)	-0.546 (71)
20. Valentine	+0.127	2.47e+05 (84)	+0.593 (15)	70. I feel	-0.173	5.17e+06 (31)	-0.129 (50)
21. Stephen Colbert	+0.126	2.38e+04 (99)	+0.001 (45)	71. man	-0.175	1.59e+07 (14)	-0.163 (52)
22. USA	+0.113	2.16e+06 (54)	+0.325 (26)	72. Republican	-0.181	2.30e+05 (86)	-0.539 (70)
23. !	+0.106	3.44e+06 (40)	+0.195 (31)	73. sad	-0.187	3.56e+06 (38)	-1.366 (89)
24. winter	+0.101	1.26e+06 (64)	+0.050 (43)	74. gas	-0.193	1.02e+06 (65)	-0.471 (67)
25. God	+0.099	8.58e+06 (25)	+0.468 (20)	75. economy	-0.203	6.09e+05 (73)	-0.525 (69)
26. hot	+0.095	7.12e+06 (28)	-0.172 (54)	76. Obama	-0.205	2.98e+06 (44)	-0.173 (55)
27. ;)	+0.094	2.61e+06 (48)	+0.326 (25)	77. Democrat	-0.226	9.32e+04 (93)	-0.384 (64)
28. Jesus	+0.094	2.03e+06 (56)	+0.247 (28)	78. Congress	-0.231	3.92e+05 (79)	-0.580 (74)
29. today	+0.092	2.56e+07 (9)	+0.126 (36)	79. hell	-0.250	6.27e+06 (30)	-1.551 (96)
30. kiss	+0.072	1.70e+06 (59)	+0.632 (11)	80. sick	-0.262	3.58e+06 (37)	-1.630 (97)
31. yes	+0.056	1.16e+07 (19)	+0.321 (27)	81. Muslim	-0.262	2.15e+05 (88)	-0.569 (73)
32. tomorrow	+0.054	1.04e+07 (21)	+0.086 (38)	82. war	-0.270	1.96e+06 (57)	-2.040 (100)
33. you	+0.052	1.73e+08 (3)	+0.111 (37)	83. Pope	-0.277	1.52e+05 (91)	-0.316 (62)
34. heaven	+0.041	7.42e+05 (71)	+0.674 (10)	84. hate	-0.282	9.65e+06 (23)	-1.520 (94)
35. ;-)	+0.041	9.39e+05 (66)	+0.395 (23)	85. Glenn Beck	-0.282	1.14e+05 (92)	-0.776 (82)
36. we	+0.035	3.91e+07 (7)	+0.146 (34)	86. Islam	-0.299	1.87e+05 (89)	-0.710 (78)
37. yesterday	+0.033	3.08e+06 (42)	-0.168 (53)	87. George Bush	-0.333	3.23e+04 (98)	-0.747 (80)
38. dark	+0.031	1.58e+06 (61)	-0.766 (81)	88. Goldman Sachs	-0.337	5.27e+04 (96)	-0.984 (84)
39. ?	+0.030	2.32e+06 (53)	-0.503 (68)	89. depressed	-0.339	2.81e+05 (82)	-1.541 (95)
40. RT	+0.028	3.39e+08 (1)	-0.443 (66)	90. Senate	-0.340	4.48e+05 (78)	-0.601 (75)
41. Michael Jackson	+0.018	8.26e+05 (70)	-0.213 (59)	91. BP	-0.355	5.82e+05 (74)	-0.902 (83)
42. night	+0.014	1.71e+07 (12)	+0.074 (40)	92. gun	-0.367	6.81e+05 (72)	-1.476 (93)
43. life	+0.012	1.40e+07 (17)	+0.422 (22)	93. drugs	-0.382	5.10e+05 (77)	-1.452 (91)
44. health	-0.000	2.58e+06 (50)	+0.447 (21)	94. headache	-0.437	8.57e+05 (69)	-1.881 (98)
45. sex	-0.008	3.55e+06 (39)	+0.542 (17)	95. :-(	-0.455	3.40e+05 (81)	-1.174 (85)
46. work	-0.010	1.84e+07 (11)	-0.174 (56)	96. :(	-0.472	2.89e+06 (45)	-1.288 (88)
47. girl	-0.010	1.01e+07 (22)	+0.331 (24)	97. Afghanistan	-0.703	2.74e+05 (83)	-1.458 (92)
48. boy	-0.026	4.93e+06 (33)	+0.062 (41)	98. mosque	-0.709	6.98e+04 (95)	-0.694 (77)
49. I	-0.048	3.08e+08 (2)	-0.062 (49)	99. flu	-0.735	9.01e+05 (68)	-1.912 (99)
50. commute	-0.048	9.01e+04 (94)	-0.206 (57)	100. Iraq	-0.773	2.39e+05 (85)	-1.282 (87)

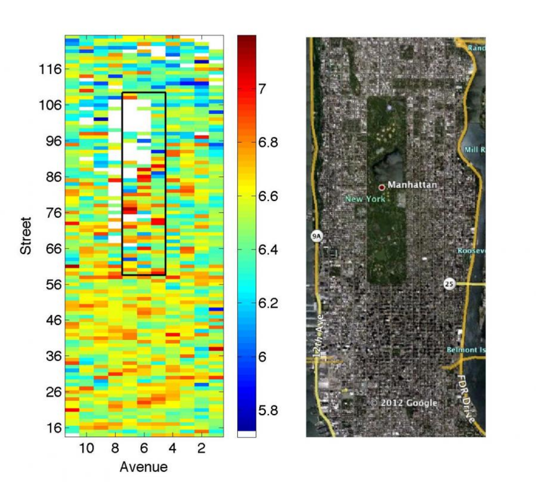
The Geography of Happiness:

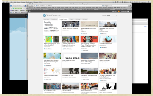


- From “The Geography of Happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place”, Mitchell et al., 2013, to appear in PLoS ONE [19].
- See blog posts here (⊞), here (⊞), and here (⊞).


40.85
40.8
40.75
40.7
40.65
40.6
40.55
-74.25
-74.2
-74.15
-74.1
-74.05
-74
-73.95
-73.9
-73.85
-73.8
-73.75
6.5
6.4
6.3
6.2
6.1
6
5.9
5.8
5.7
5.6
5.5

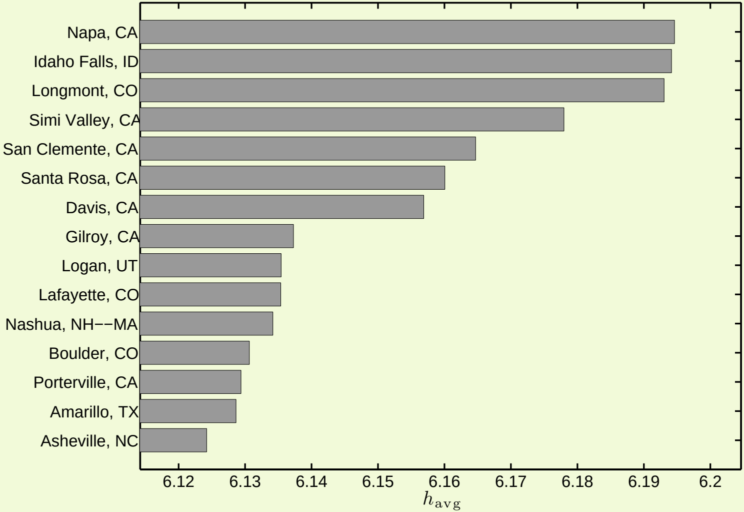
Happiness in Manhattan:



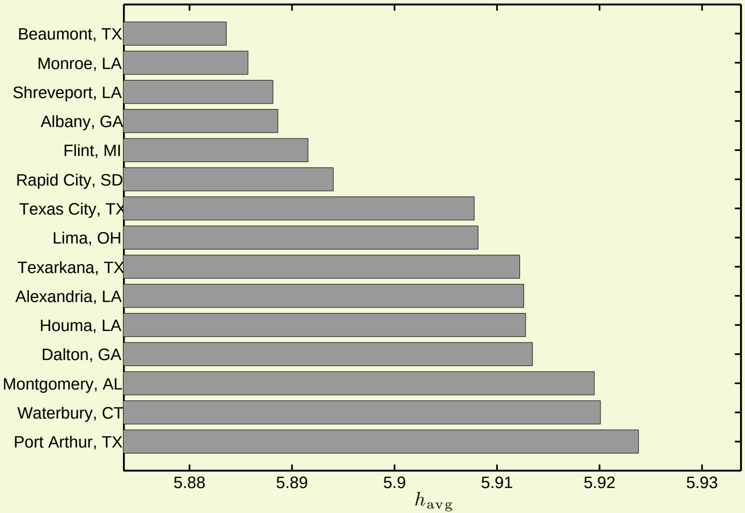


See Blog post on onehappybird (⊞)

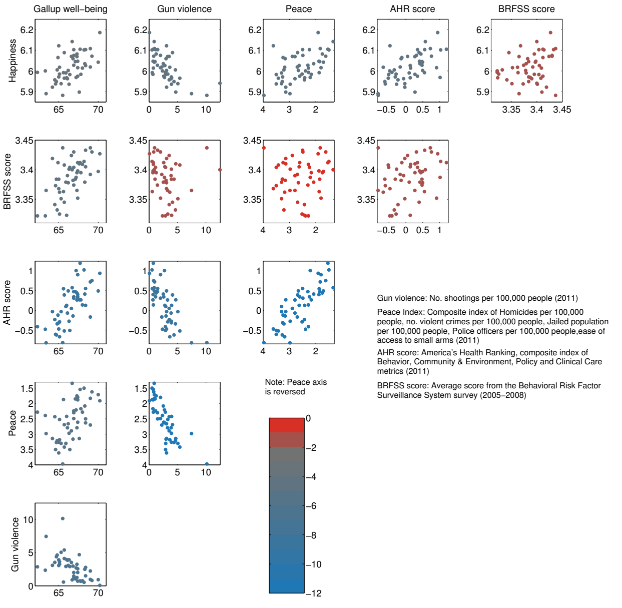
Happiest Cities:



Saddest Cities (geoprofanity):

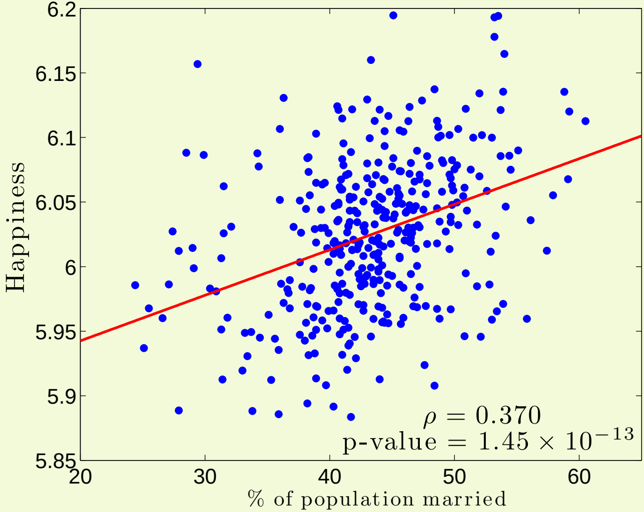


Validity test #30,231(b):

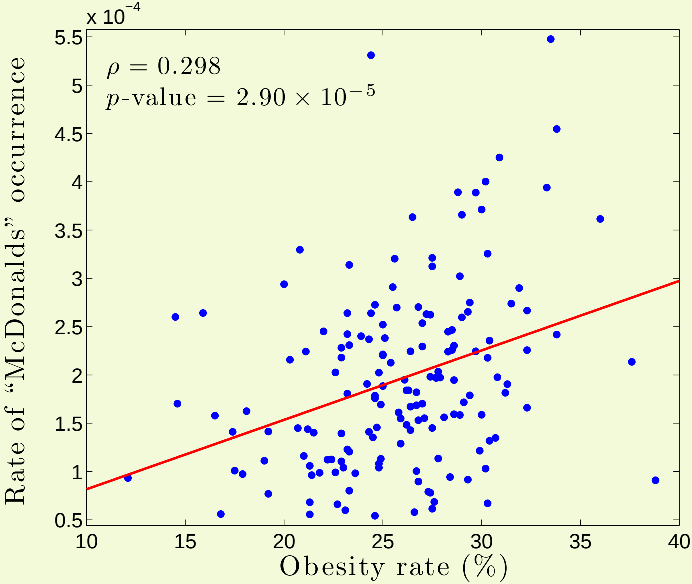


Good news for Valentine's Day:

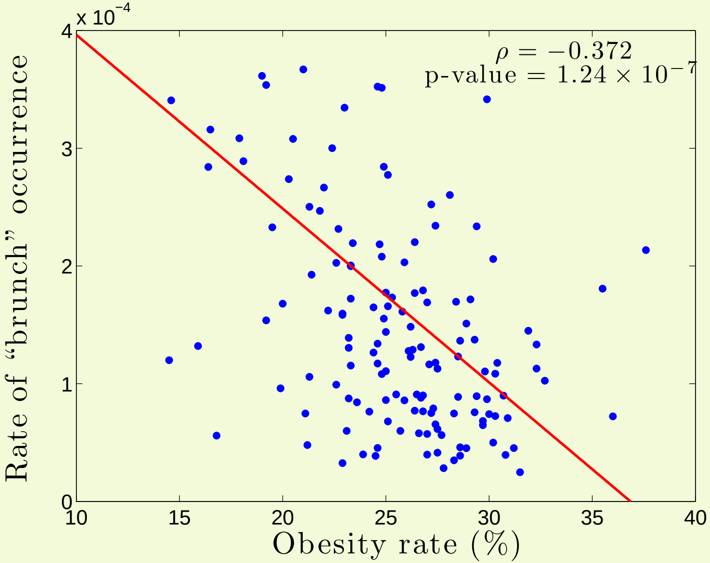
Happiness and Marriage:



Obesity and tweets—“McDonalds”:



Obesity and tweets—“Brunch”:



Obesity rates and usage of food-related words:

Negative correlations

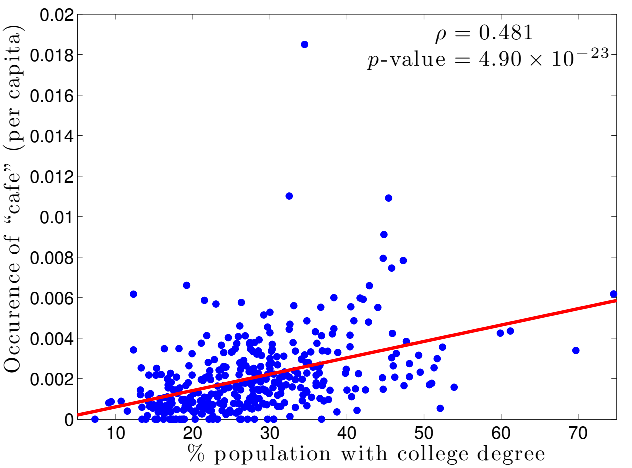
Word	ρ	p-value
cafe	-0.509	6.07×10^{-14}
sushi	-0.487	9.93×10^{-13}
brewery	-0.469	8.67×10^{-12}
restaurant	-0.448	8.93×10^{-11}
bar	-0.435	3.59×10^{-10}
banana	-0.434	3.77×10^{-10}
apple	-0.408	5.22×10^{-9}
fondue	-0.403	8.34×10^{-9}
wine	-0.400	1.08×10^{-8}
delicious	-0.392	2.17×10^{-8}
dinner	-0.386	3.85×10^{-8}
coffee	-0.384	4.51×10^{-8}
bakery	-0.383	5.12×10^{-8}
bean	-0.378	7.88×10^{-8}
espresso	-0.377	8.47×10^{-8}
cuisine	-0.376	8.82×10^{-8}
foods	-0.374	1.07×10^{-7}
tofu	-0.372	1.27×10^{-7}
brunch	-0.368	1.79×10^{-7}
veggie	-0.364	2.46×10^{-7}
organic	-0.361	3.13×10^{-7}
booze	-0.360	3.34×10^{-7}
grill	-0.354	5.4×10^{-7}
chocolate	-0.351	6.77×10^{-7}
#vegan	-0.350	7.47×10^{-7}

Obesity rates and usage of food-related words:

Positive correlations

mcdonalds	0.246	6.18×10^{-4}
eat	0.241	8.22×10^{-4}
wings	0.222	2.13×10^{-3}
hungry	0.210	3.65×10^{-3}
heartburn	0.194	7.37×10^{-3}
ham	0.177	1.45×10^{-2}

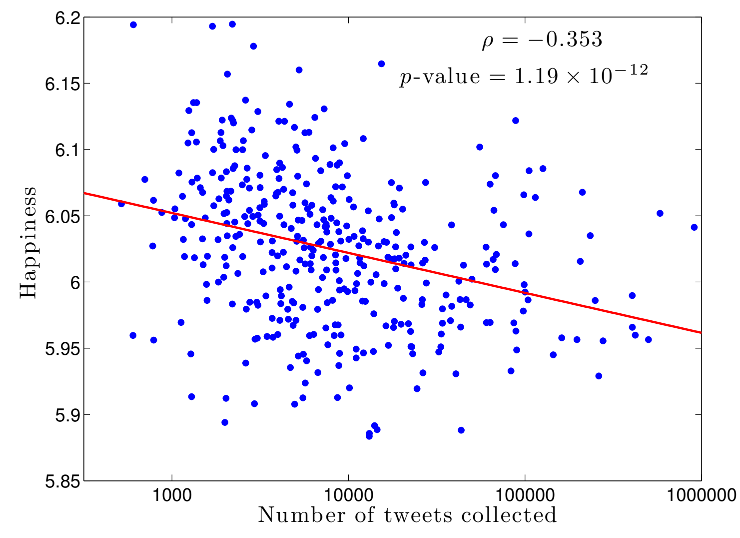
‘cafe’ usage frequency vs. fraction with College degree:

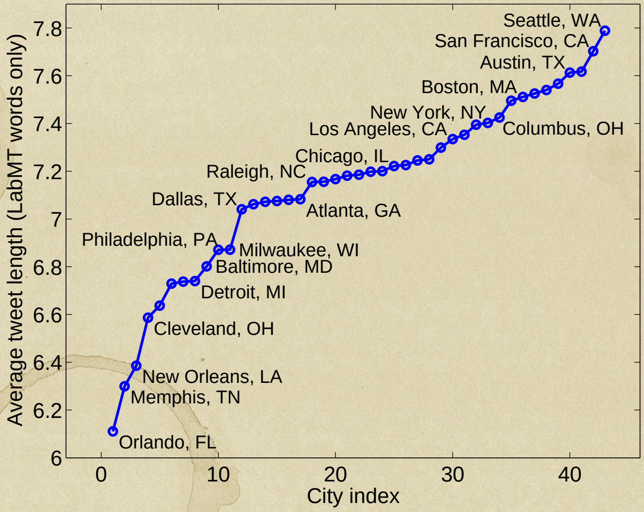


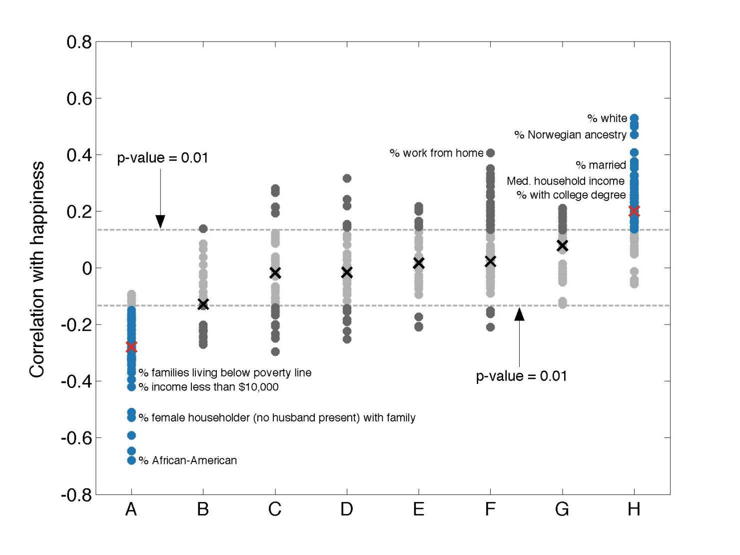
Word usage frequency vs. fraction with College degree:

Word	ρ	p-value	$h_{\text{avg}}(w_i)$
cafe	0.481	4.9×10^{-23}	6.78
pub	0.463	3.14×10^{-21}	6.02
software	0.458	9.07×10^{-21}	6.30
yoga	0.455	1.85×10^{-20}	7.04
grill	0.433	1.78×10^{-18}	6.24
development	0.424	1.14×10^{-17}	6.38
emails	0.419	2.87×10^{-17}	6.54
wine	0.417	3.83×10^{-17}	6.42
library	0.414	6.47×10^{-17}	6.48
art	0.414	6.8×10^{-17}	6.60
sciences	0.410	1.54×10^{-16}	6.30
pasta	0.410	1.57×10^{-16}	6.86
lounge	0.409	1.68×10^{-16}	6.50
market	0.408	2.2×10^{-16}	6.28
india	0.407	2.5×10^{-16}	6.42
drinking	0.405	3.74×10^{-16}	6.14
technology	0.405	3.76×10^{-16}	6.74
forest	0.405	3.83×10^{-16}	6.68
brunch	0.405	3.89×10^{-16}	6.32
dining	0.403	4.92×10^{-16}	6.48
supporting	0.399	1.1×10^{-15}	6.48
professor	0.398	1.23×10^{-15}	6.04
university	0.392	3.62×10^{-15}	6.74
film	0.391	4.27×10^{-15}	6.56
global	0.391	4.72×10^{-15}	6.00

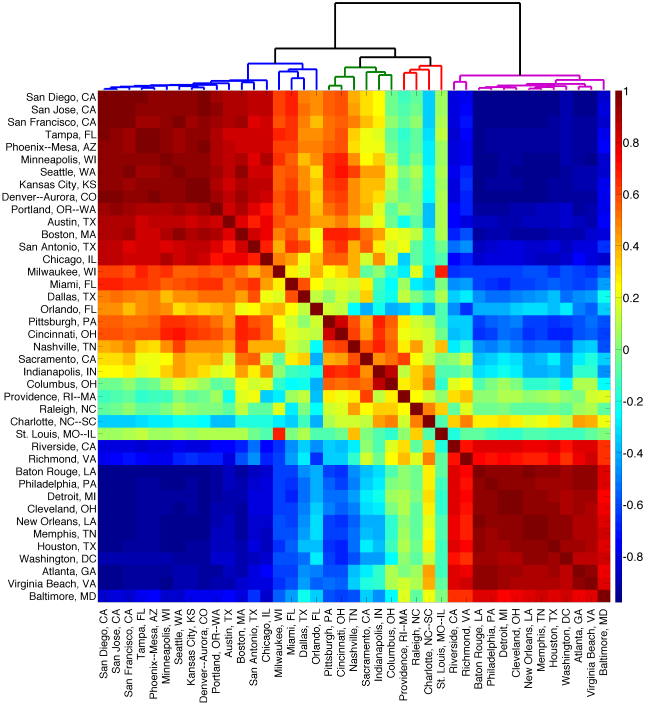
Word	ρ	p-value	$h_{\text{avg}}(w_i)$
me	-0.393	3.26×10^{-15}	6.58
love	-0.389	6.51×10^{-15}	8.42
my	-0.354	1.97×10^{-12}	6.16
like	-0.346	6.04×10^{-12}	7.22
hate	-0.344	8.76×10^{-12}	2.34
tired	-0.343	1×10^{-11}	3.34
sleep	-0.341	1.27×10^{-11}	7.16
stupid	-0.328	8.55×10^{-11}	2.68
bored	-0.315	5.11×10^{-10}	3.04
you	-0.315	5.23×10^{-10}	6.24
goodnight	-0.305	1.77×10^{-9}	6.58
bitch	-0.295	6.51×10^{-9}	3.14
all	-0.289	1.33×10^{-8}	6.22
lie	-0.285	2.24×10^{-8}	2.60
mom	-0.284	2.42×10^{-8}	7.64
wish	-0.271	1.05×10^{-7}	6.92
talk	-0.267	1.74×10^{-7}	6.06
she	-0.265	2.01×10^{-7}	6.18
know	-0.262	2.78×10^{-7}	6.10
ill	-0.259	4.11×10^{-7}	2.42
dont	-0.258	4.54×10^{-7}	3.70
well	-0.256	5.3×10^{-7}	6.68
don't	-0.255	5.8×10^{-7}	3.70
give	-0.255	5.84×10^{-7}	6.54
friend	-0.255	6.27×10^{-7}	7.66


$\rho = -0.353$
p-value $= 1.19 \times 10^{-12}$
6.2
6.15
6.1
6.05
6
5.95
5.9
5.85
Happiness
1000
10000
100000
1000000
Number of tweets collected


Average tweet length (LabMT words only)
City index
Seattle, WA
San Francisco, CA
Austin, TX
Boston, MA
New York, NY
Columbus, OH
Los Angeles, CA
Chicago, IL
Raleigh, NC
Dallas, TX
Atlanta, GA
Philadelphia, PA
Milwaukee, WI
Baltimore, MD
Detroit, MI
Cleveland, OH
New Orleans, LA
Memphis, TN
Orlando, FL
7.8
7.6
7.4
7.2
7
6.8
6.6
6.4
6.2
6
0
10
20
30
40



Explore more here (⊞).



Finding Happiness

- Measuring Happiness
- Data sets
  - Blogs
- Twitter
- Geography
- Health
- **Demographics**
- Movement
- Networks
- Phrases
- The End
- References

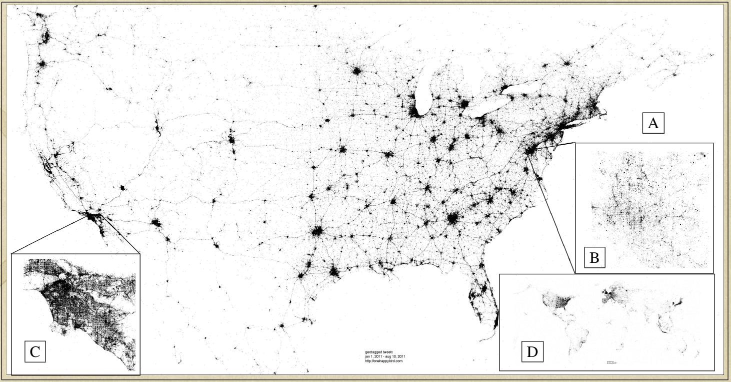
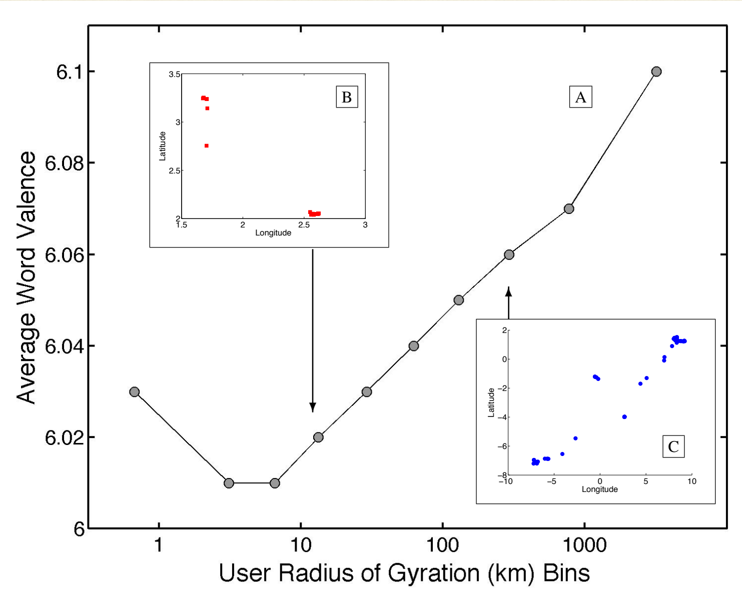
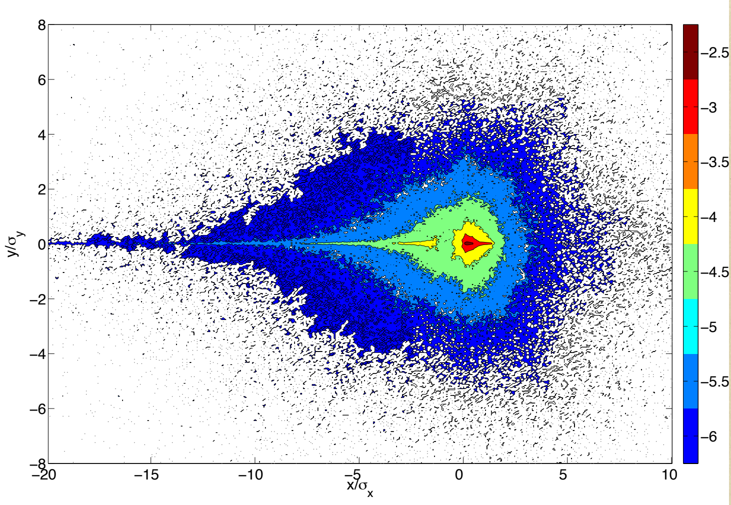


Figure 1. Each point corresponds to a geo-located tweet posted between 1/1/11 and 8/10/11. Twitter activity seems to correlate with urban areas. Note that the image contains no cartographic borders, simply a small dot for each message. Insets: **A** (U.S.), **B** (Washington, D.C.), **C** (Los Angeles, C.A.), and **D** (Earth).

- From "Happiness and the Patterns of Life: A Study of Geolocated Tweets", Frank et al., 2013, in review [9].
- See blog post here (⊞).

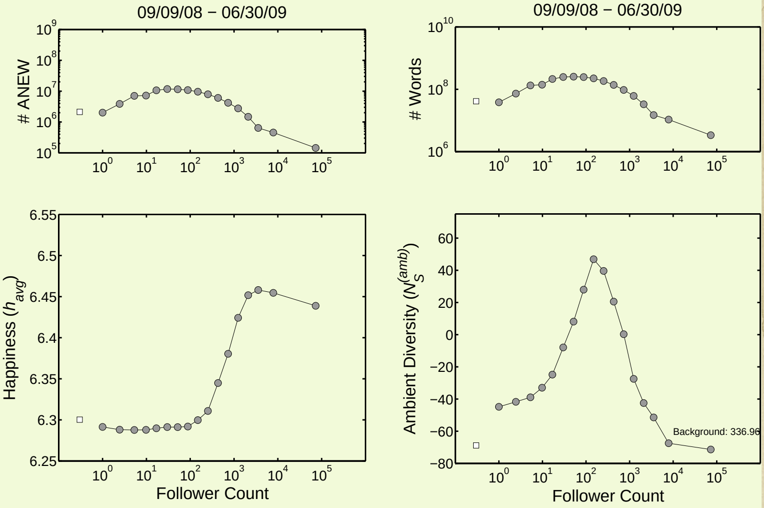


Frank et al., in preparation.

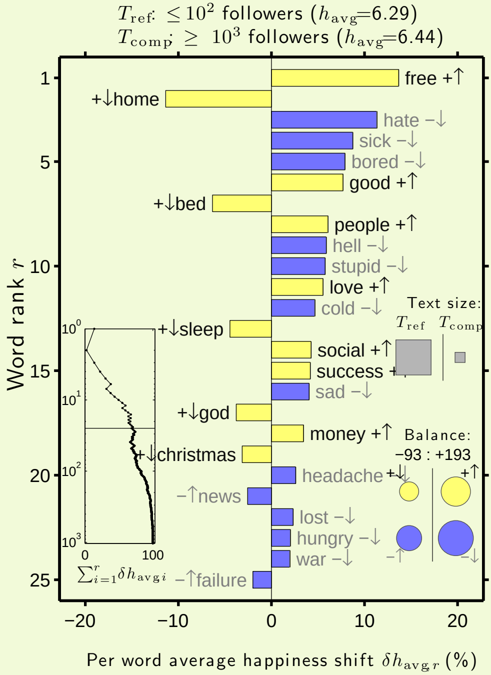


Raw movement patterns agree with cell phone data findings [10]

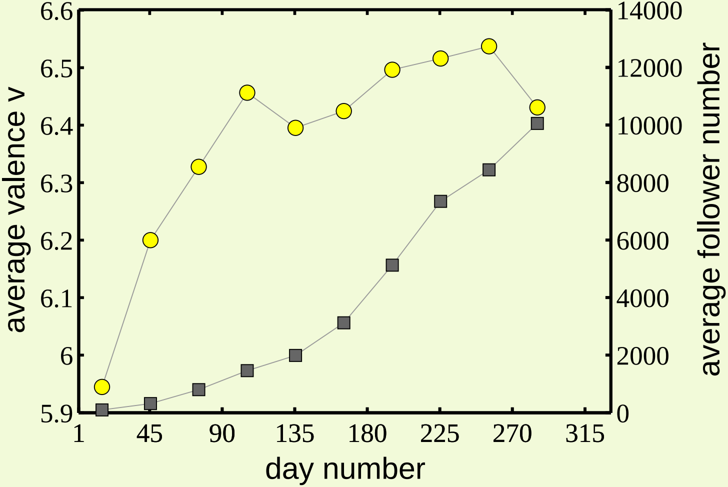
Twitter—popularity based on follower count:



- Dunbar's number $\simeq$ 150.

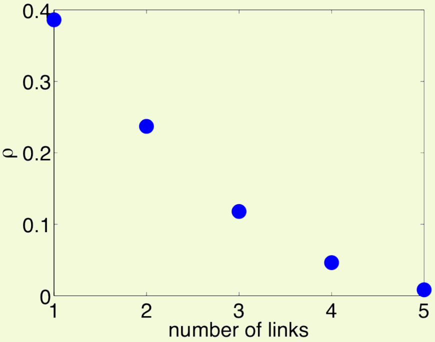


$T_{\rm ref}$: $\leq 10^2$ followers ($h_{\rm avg}$=6.29)
$T_{\rm comp}$: $\geq 10^3$ followers ($h_{\rm avg}$=6.44)
free +↑
+↓home
hate −↓
sick −↓
bored −↓
good +↑
+↓bed
people +↑
hell −↓
stupid −↓
love +↑
cold −↓
+↓sleep
social +↑
success +↑
sad −↓
+↓god
money +↑
+↓christmas
headache +↓
−↑news
lost −↓
hungry −↓
war −↓
−↑failure
Text size:
$T_{\rm ref}$ $T_{\rm comp}$
Balance:
−93 : +193
+↓ +↑
−↑ −↓
Word rank r
1
5
10
15
20
25
10^0
10^1
10^2
10^3
0
100
$\sum_{i=1}^{r} \delta h_{{\rm avg}\,i}$
−20
−10
0
10
20
Per word average happiness shift $\delta h_{{\rm avg},r}$ (%)



6.6
6.5
6.4
6.3
6.2
6.1
6
5.9
average valence v
14000
12000
10000
8000
6000
4000
2000
0
average follower number
1
45
90
135
180
225
270
315
day number

Twitter—interactions:



- Decay in happiness correlation in social network.
- ρ = Spearman's correlation coefficient.
- "Twitter reciprocal reply networks exhibit assortativity with respect to happiness"
  Bliss et al., Journal of Computational Science, 2012 [1]

Phrases—Music Lyrics:

rank	order=1	order=2	order=3	order=4
1	i	and i	i know you	if you want to
2	the	in the	you know i	let me tell you
3	and	if you	and i know	tell me what you
4	you	on the	this is then	don’t want to be
5	a	to the	la la la	all i need is
6	to	i know	don’t want to	and i know that
7	my	you know	if i could	what can i do
8	i’m	but i	can’t you see	want you to know
9	it	when i	don’t know what	all i want is
10	that	when you	all the time	give it to me
11	so	all the	why don’t you	when it comes to
12	your	like a	as long as	how does it feel
13	me	this is	don’t you know	you know that i
14	in	come on	there is no	don’t you know that
15	no	to be	i know that	don’t give a fuck
25	love	don’t know	but i can’t	all the things that
100	m	just like	in this world	woke up this morning

- J. Williams et al., in preparation.

Next for Happiness:

- hedonometer.org (⊞) (early 2013).
- Over 10 additional languages being scored through a new service.
- Four other emotions: surprise, fear, disgust, and anger.
- Other input streams (e.g., BBC)
- Expansion to phrase-based analysis.

"Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter"
Dodds et al., PLoS ONE, 2011 [8]
Much better version here:
http://arxiv.org/abs/1101.5120 (⊞)

- "Twitter reciprocal reply networks exhibit assortativity with respect to happiness"
  Bliss et al., Journal of Computational Science, 2012 [1]
- "Positivity of the English Language"
  Kloumann et al., PLoS ONE, 2012 [14]
- "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents"
  Dodds and Danforth, Journal of Happiness Studies, 2009 [7]
- language assessment by Mechanical Turk
  (labMT 1.0)
- http://www.onehappybird.com (⊞)

Some press...

- "Social Scientists wade into the Tweet stream" by Greg Miller, Science, **333**, 1814–1815, 2011 [18]
- "Does a Nation's Mood Lurk in Its Songs and Blogs?" by Benedict Carey New York Times, August 2009. (⊞)
- More here: http://www.uvm.edu/∼pdodds/research/ (⊞)

References I

[1] C. A. Bliss, I. M. Kloumann, K. D. Harris, C. M. Danforth, and P. S. Dodds.
Twitter reciprocal reply networks exhibit assortativity with respect to happiness.
Journal of Computational Science, 3:388–397, 2012. pdf (⊞)

[2] M. Bradley and P. Lang.
Affective norms for english words (anew): Stimuli, instruction manual and affective ratings.
Technical report c-1, University of Florida, Gainesville, FL, 1999. pdf (⊞)

References II

[3] T. Conner Christensen, L. Feldman Barrett, E. Bliss-Moreau, K. Lebo, and C. Kaschub.
A practical guide to experience-sampling procedures.
Journal of Happiness Studies, 4:53–78, 2003.

[4] M. Csikszentmihalyi.
Flow.
Harper & Row, New York, 1990.

[5] M. Csikszentmihalyi, R. Larson, and S. Prescott.
The ecology of adolescent activity and experience.
Journal of Youth and Adolescence, 6:281–294, 1977.

References III

[6] E. Diener and M. Y. Chan.
Happy people live longer: Subjective well-being contributes to health and longevity.
Applied Psychhology: Health and Well-Being, 3:1–43, 2011. pdf (⊞)

[7] P. S. Dodds and C. M. Danforth.
Measuring the happiness of large-scale written expression: Songs, blogs, and presidents.
Journal of Happiness Studies, 2009. doi:10.1007/s10902-009-9150-9. pdf (⊞)

[8] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth.
Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter.
PLoS ONE, 6:e26752, 2011.

References IV

Draft version available at http://arxiv.org/abs/1101.5120v4. Accessed October 24, 2011. pdf (⊞)

[9] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth.
Happiness and the patterns of life: A study of geolocated Tweets.
http://arxiv.org/abs/1304.1296, 2013. pdf (⊞)

[10] M. C. González, C. A. Hidalgo, and A.-L. Barabási.
Understanding individual human mobility patterns.
Nature, 453:779–782, 2008. pdf (⊞)

[11] W. T. Jones.
The Classical Mind.
Harcourt, Brace, Jovanovich, New York, 1970.

References V

[12] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone.
A survey method for characterizing daily life experience: The day reconstruction method.
Science, 306(5702):1776–1780, 2004. pdf (⊞)

[13] M. A. Killingsworth and D. T. Gilbert.
A wondering mind is an unhappy mind.
Science Magazine, 330:932, 2010. pdf (⊞)

[14] I. M. Kloumann, C. M. Danforth, K. D. Harris, C. A. Bliss, and P. S. Dodds.
Positivity of the English language.
PLoS ONE, 7:e29484, 2012.
Draft version available at http://arxiv.org/abs/1108.5192. Accessed October 24, 2011. pdf (⊞)

References VI

[15] R. Layard.
Happiness.
The Penguin Press, London, 2005.

[16] S. Lyubomirsky.
The How of Happiness.
The Penguin Press, New York, 2007.

[17] C. Martinelli and S. W. Parker.
Deception and misreporting in a social program.
forthcoming in Journal of the European Economic Association, 2007. pdf (⊞)

[18] G. Miller.
Social Scientists wade into the Tweet stream.
Science Magazine, 333:1814–1815, 2011. pdf (⊞)

References VII

[19] L. Mitchell, M. R. Frank, P. S. Dodds, and C. M. Danforth.
The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place.
http://arxiv.org/abs/1302.3299, 2013.
pdf (⊞)

[20] C. Osgood, G. Suci, and P. Tannenbaum.
The Measurement of Meaning.
University of Illinois, Urbana, IL, 1957.