Measuring Happiness the Big Data Way

Clinical and Translational Research Seminar, UVM

Peter Dodds, Chris Danforth,
Isabel Kloumann, Cathy Bliss, and Kameron Harris.

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

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The UNIVERSITY of VERMONT

Outline

Some motivation

Measuring emotional content

Data sets

Analysis
- Songs
- Blogs
- SOTU
- Twitter

Mechanical Turk action

Prediction

References

Some motivation

Measuring emotional content

Data sets

Analysis
- Songs
- Blogs
- SOTU
- Twitter

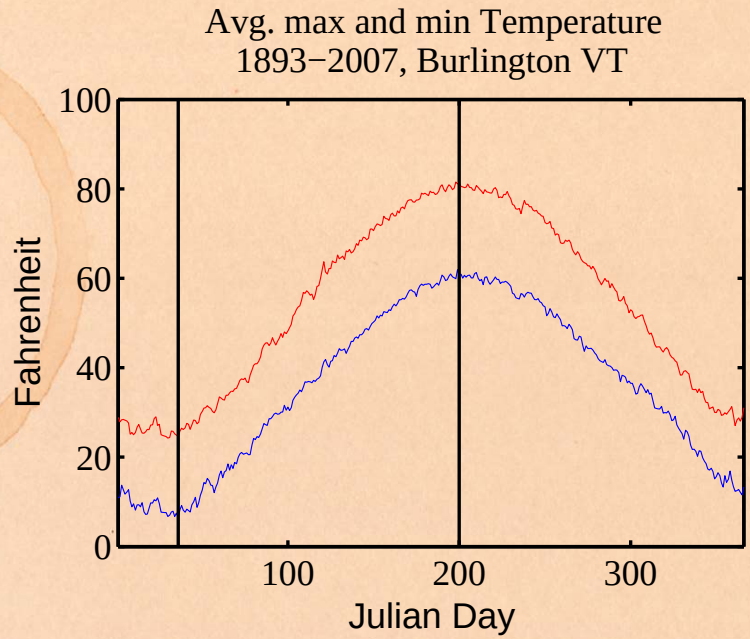
Mechanical Turk action

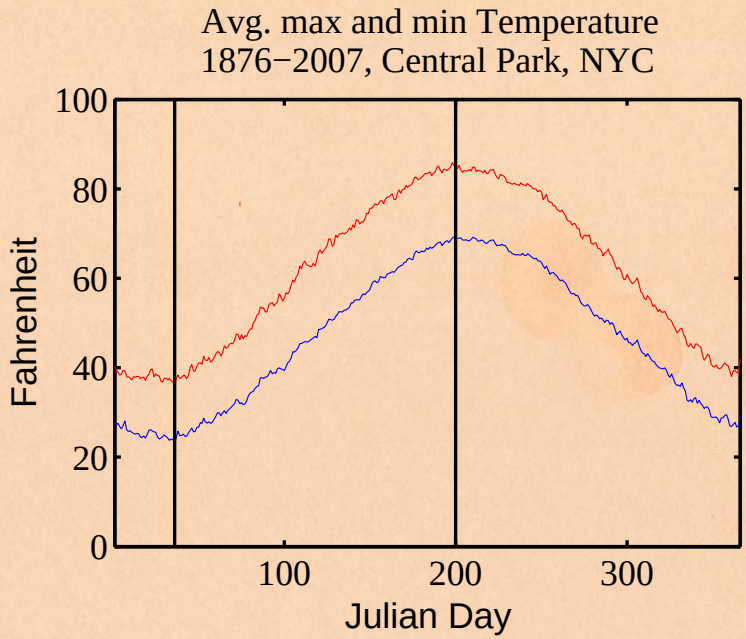
Prediction

References

The UNIVERSITY of VERMONT

Tonight is the on-average coldest night of the year:





- Hibernal Teletherm ≈ February 4.
- Halfway between Winter Solstice and Spring Equinox
- Bonus: Groundhog Day (⊞), Imbolc (⊞), . . .
- Aesteval Teletherm ≈ July 19 (164 days later).

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The Team:

The People:

Kameron Harris

Isabel Kloumann

Catherine Bliss

The Machines:

- 1400 processors + storage at the Vermont Advanced Computing Center
- 30 TB of storage in Danforth's office.

Happiness

Some motivation

Measuring emotional content

Data sets

Analysis

Songs

Blogs

SOTU

Twitter

Mechanical Turk action

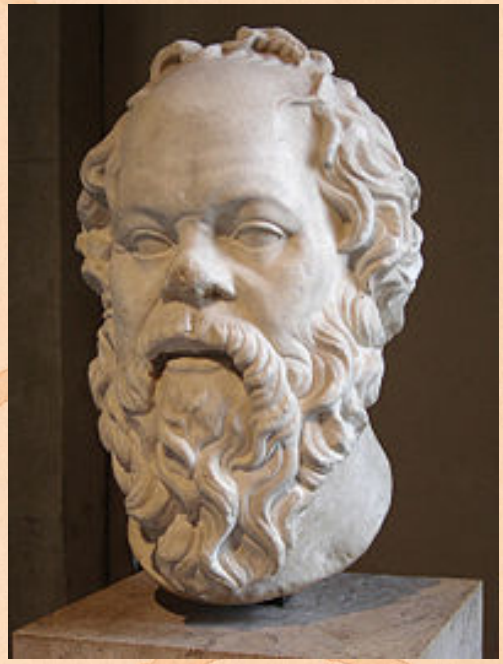
Prediction

References

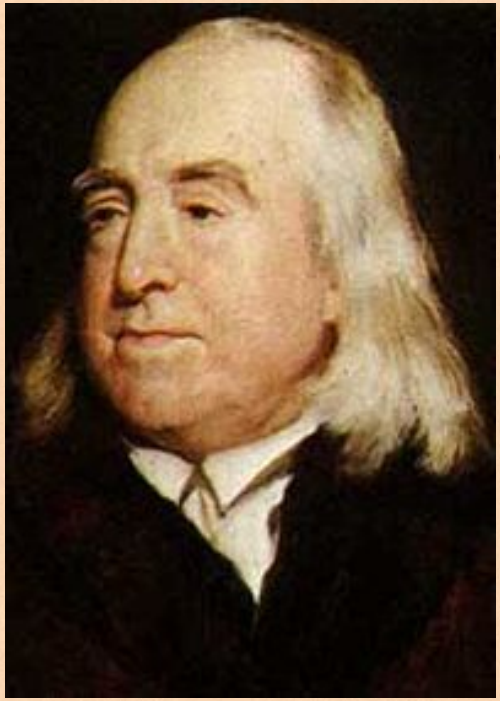
The UNIVERSITY of VERMONT

Happiness:

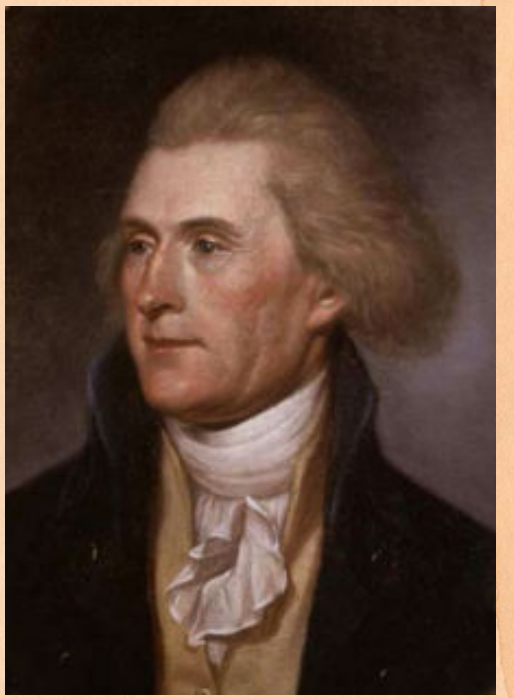
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



Socrates et al.: eudaimonia [8]

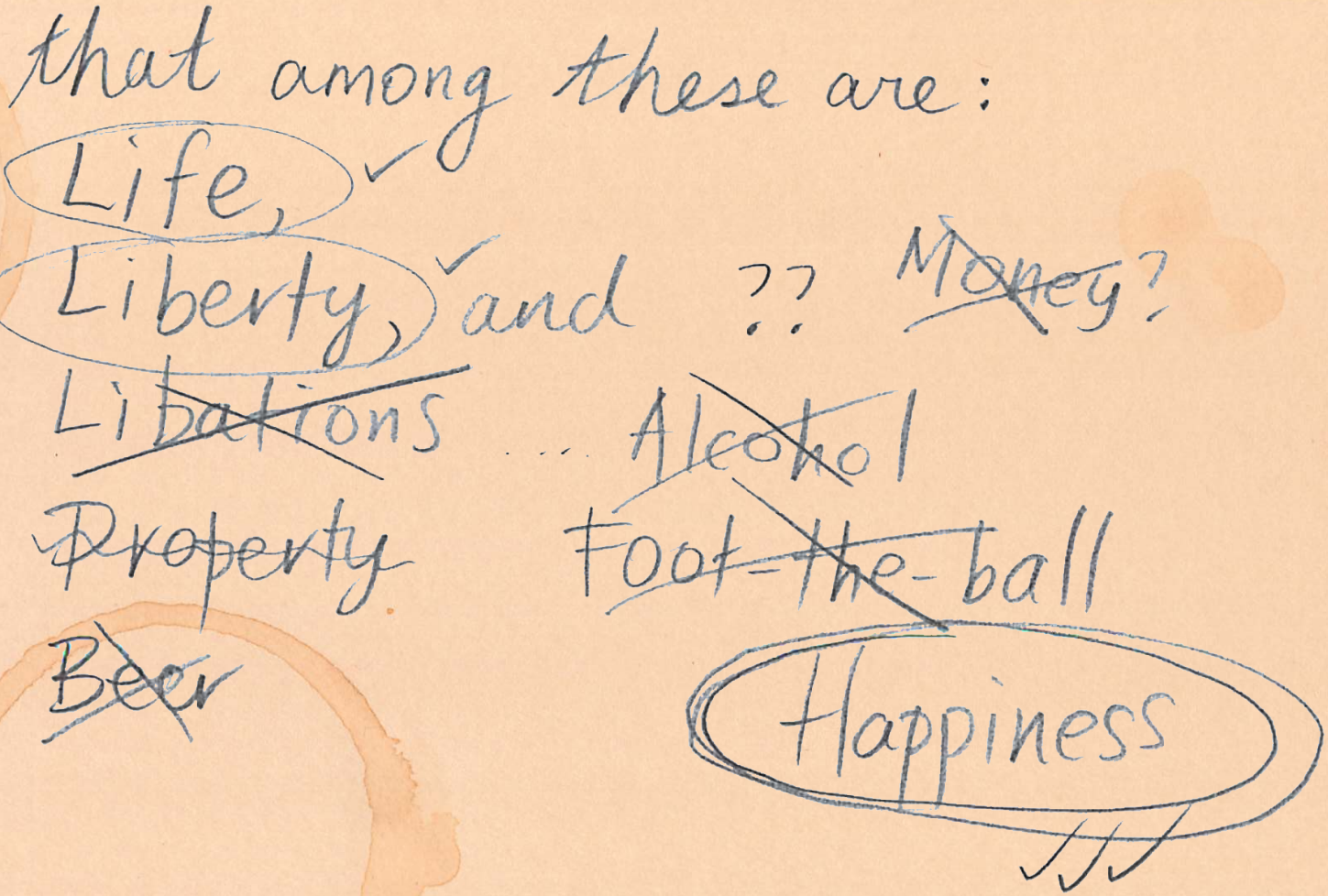


Bentham: hedonistic calculus



Jefferson: . . . the pursuit of happiness

Early drafts:



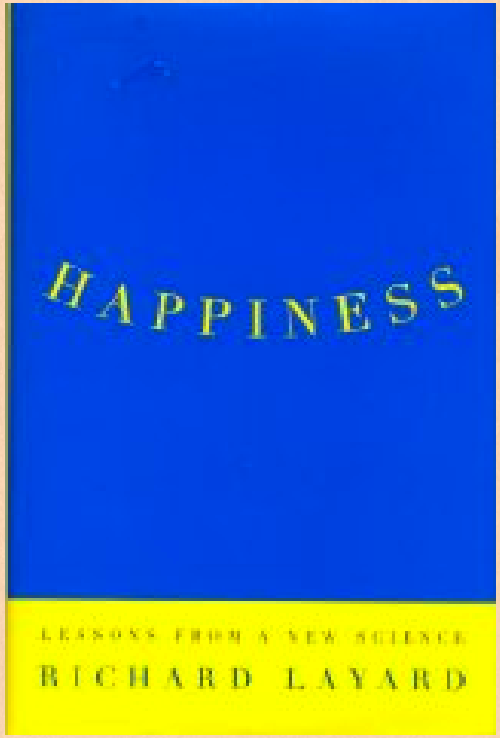
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Happiness:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Even the odd modern economist likes happiness:

"Happiness" by Richard Layard [10]



[amazon] (⊞)

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Desiring happiness—not just for boffins:

- Average people routinely report being happy is what they want most in life [10, 11]

National indices of well-being:

- Bhutan
- France
- Australia

Emotional content

So how does one measure

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

Just ask people how happy they are.

- Experience sampling [4, 6, 5] (Csikszentmihalyi et al.)
- Day reconstruction [9] (Kahneman et al.)

But self-reporting has drawbacks...

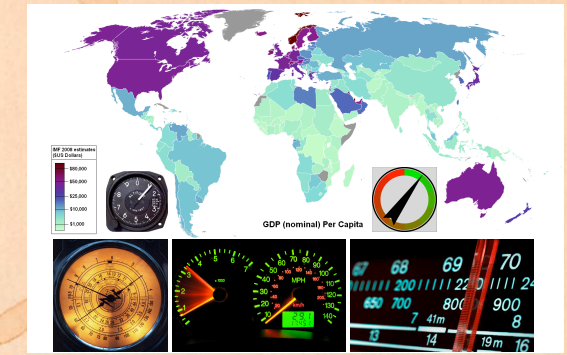
- relies on memory and self-perception
- induces misreporting [12]
- costly

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Measuring Emotional Content:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

We'd like to build an 'hedonometer':

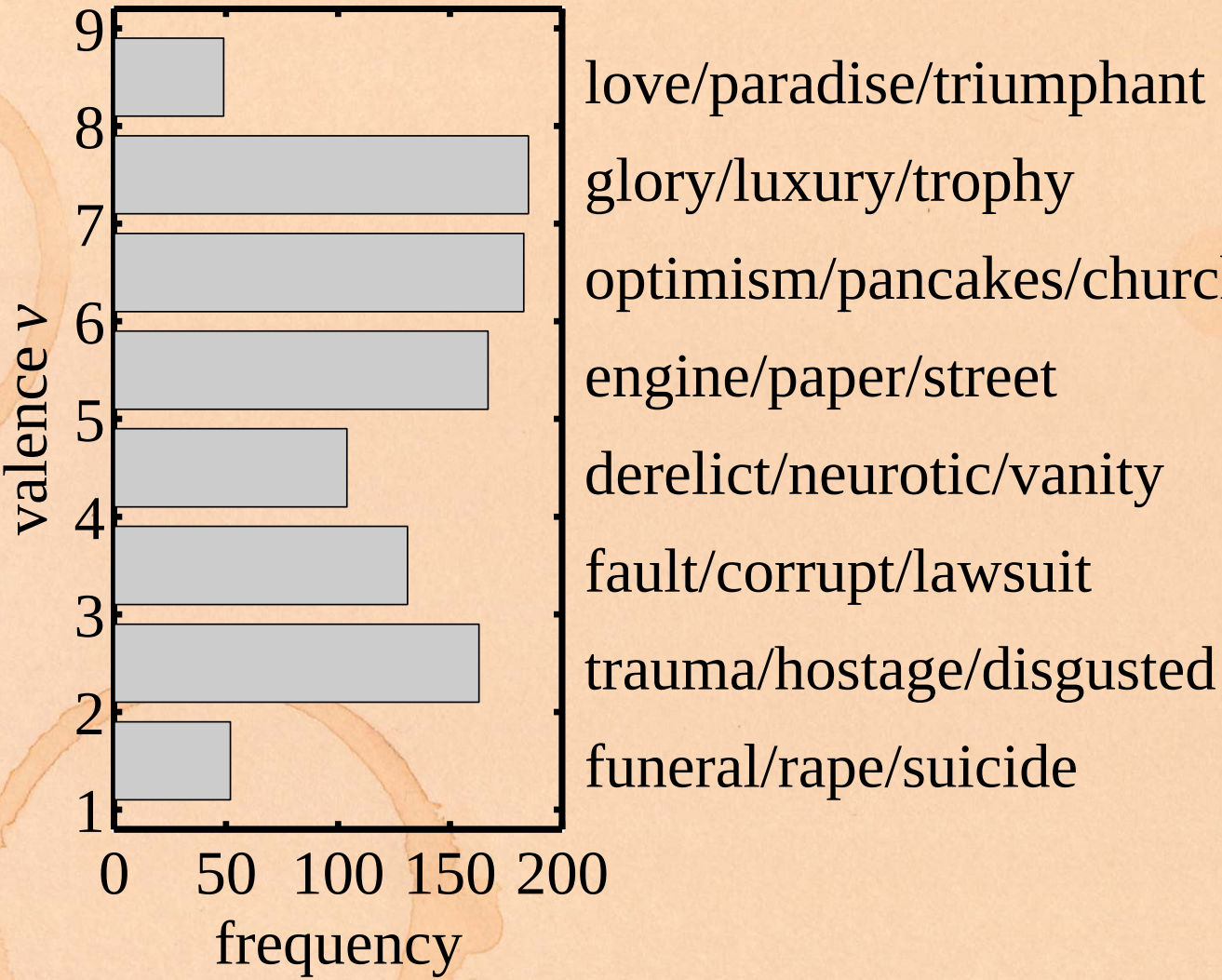


- An instrument to 'remotely-sense' emotional states and levels, in real time or post hoc.

Ideally:

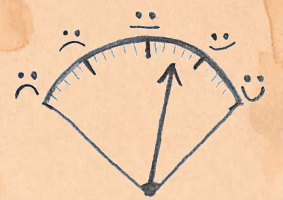
- Transparent
- Fast
- Based on written expression
- Uses human evaluation
- Non-reactive
- Complementary to self-reported measures
- Improvable

ANEW study words—examples



ANEW = “Affective Norms for English Words” [3]

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



Analysing text:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

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.
⋮

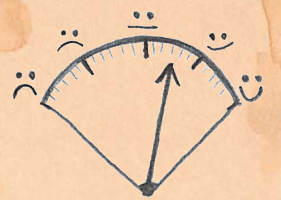
ANEW words	v_k	f_k
k=1. love	8.72	1
2. mother	8.39	1
3. baby	8.22	3
4. beauty	7.82	1
5. truth	7.80	1
6. people	7.33	2
7. strong	7.11	1
8. young	6.89	2
9. girl	6.87	4
10. movie	6.86	1
11. perfume	6.76	1
12. queen	6.44	1
13. name	5.55	1
14. lie	2.79	1

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

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

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

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



Data sets:

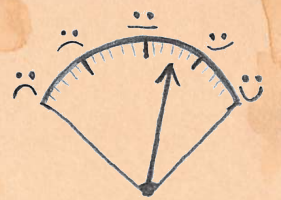
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Texts:

1. Song lyrics (1960–2007)
2. Song titles (1960–2008)
3. State of the Union (SOTU) Addresses (1790–2008)

Sources:

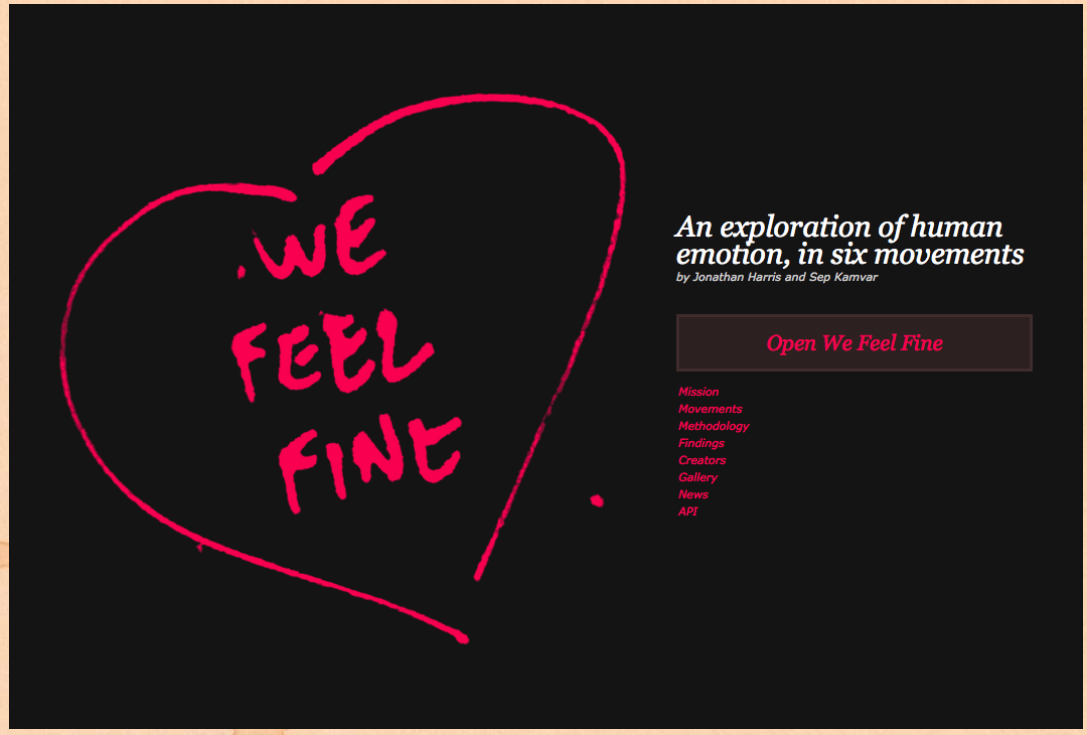
- hotlyrics.com (⊞)
- freedb.com (⊞)
- American Presidency Project: www.presidency.ucsb.edu (⊞).



The UNIVERSITY of VERMONT

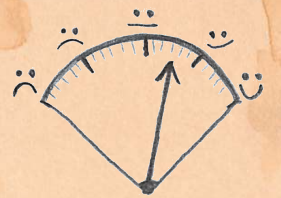
Data sets:

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



- Created by Jonathan Harris & Sep Kamvar

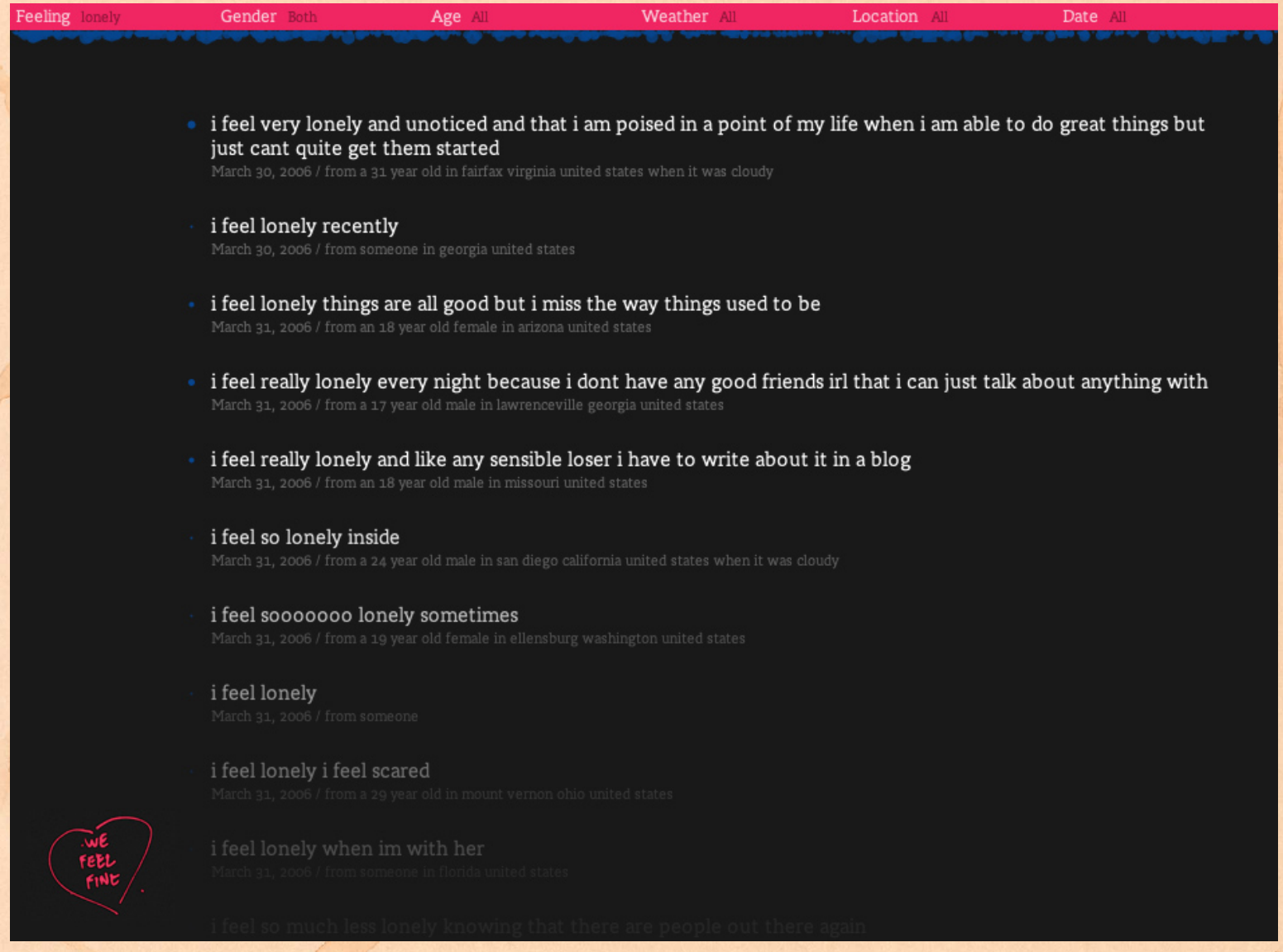
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

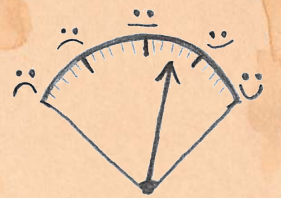


The UNIVERSITY of VERMONT

wefeelfine.org:

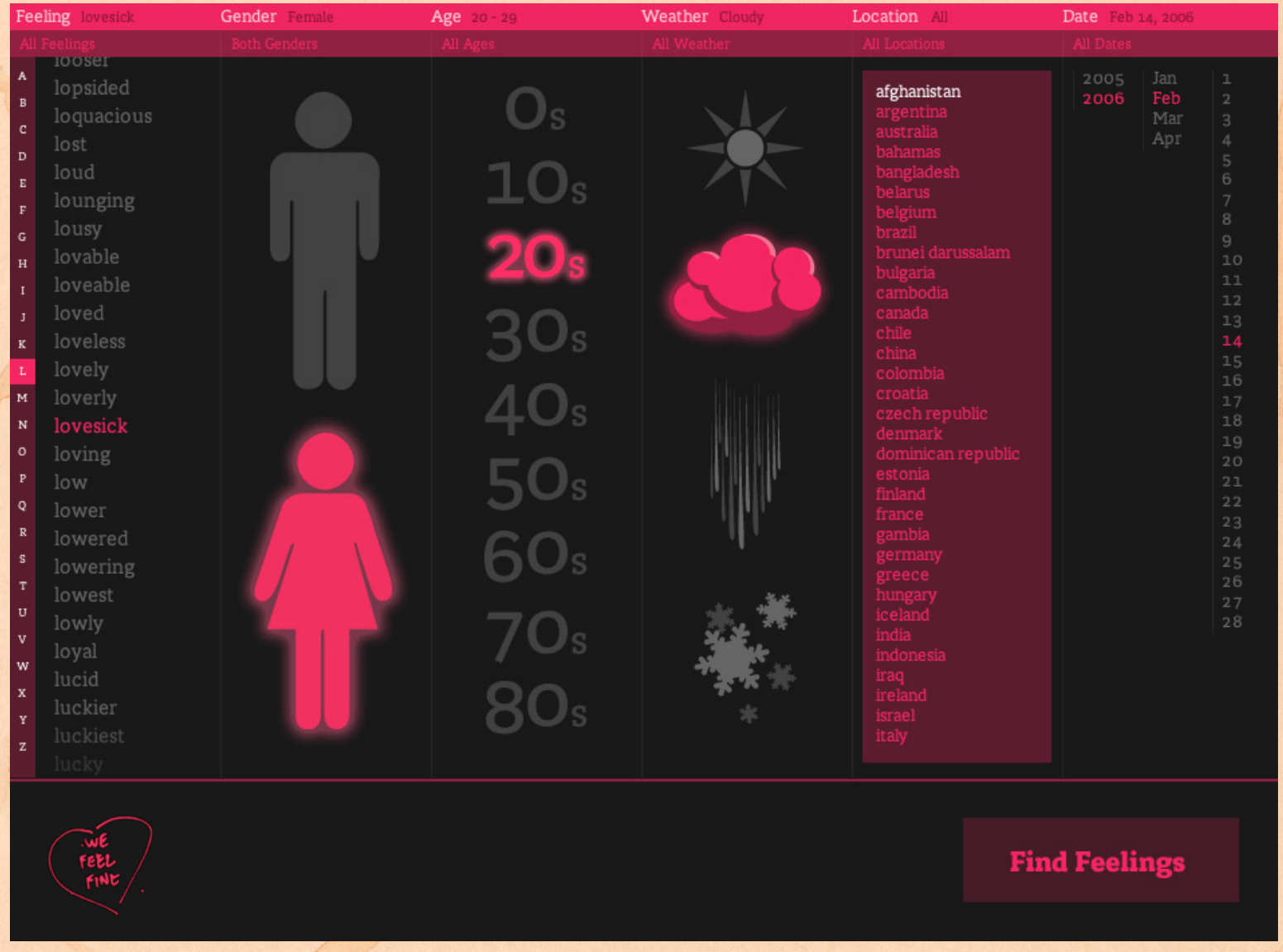
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



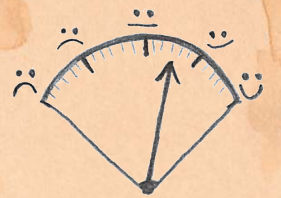


The University of Vermont

wefeelfine.org:

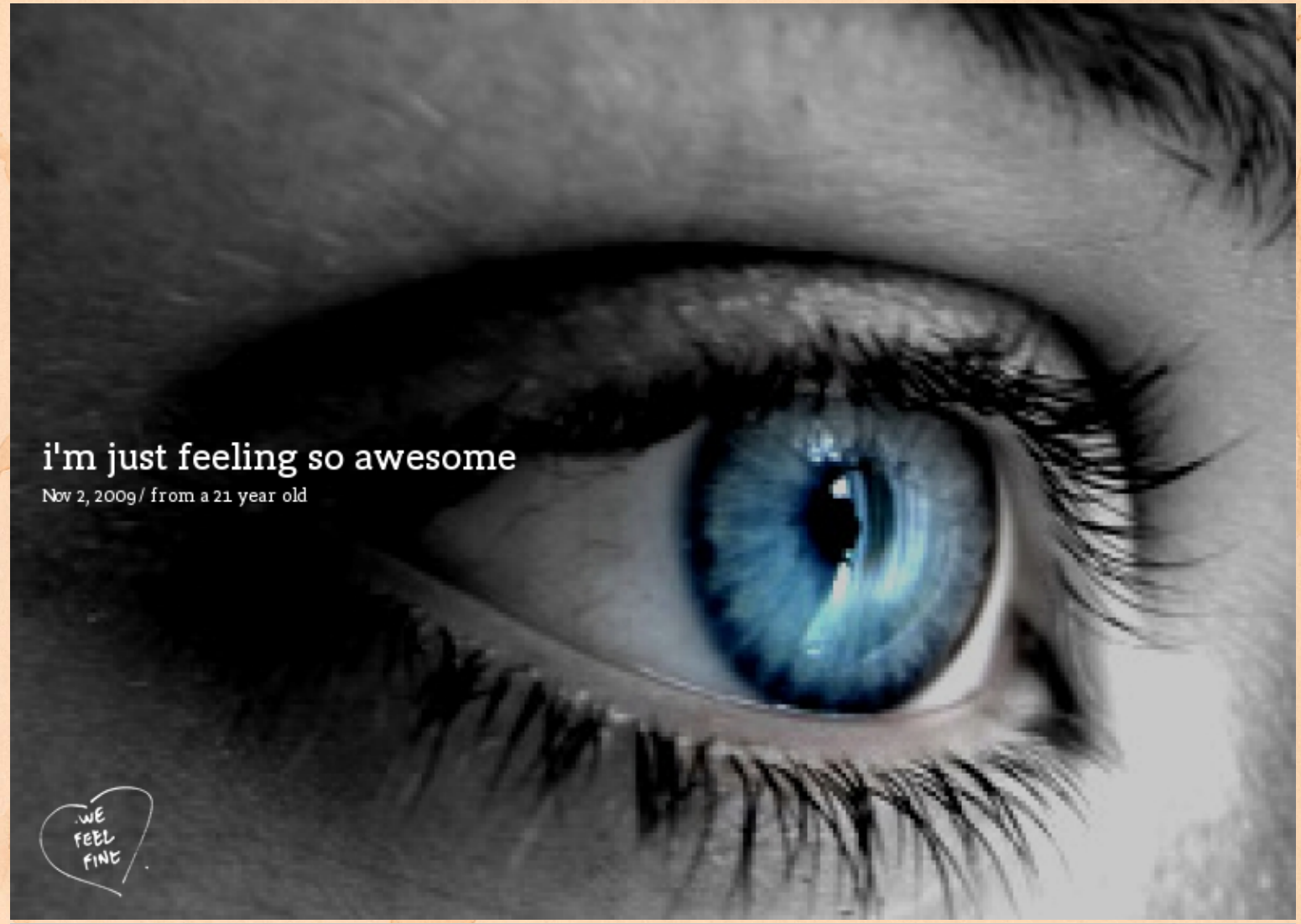


Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

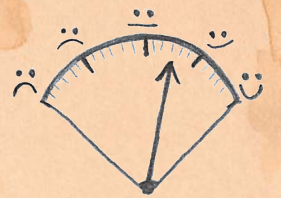


The University of Vermont

wefeelfine.org:



Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

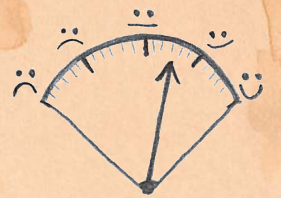


The UNIVERSITY of VERMONT

Data sets:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

5. Tweets from twitter.com:



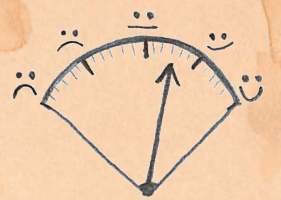
Some numbers:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Counts	Song lyrics	Song titles
All words	58,610,849	60,867,223
ANEW words	3,477,575 (5.9%)	5,612,708 (9.2%)
Individuals	$\sim$ 20,000	$\sim$ 632,000

Counts	blogs	SOTU
All words	155,667,394	1,796,763
ANEW words	8,581,226 (5.5%)	61,926 (3.5%)
Individuals	$\sim$ 2,335,000	43

Counts	Twitter
All words	$\sim 30 \times 10^9$
ANEW words	$\sim 1 \times 10^9$ (3.7%)
Individuals	$\sim 50 \times 10^6$



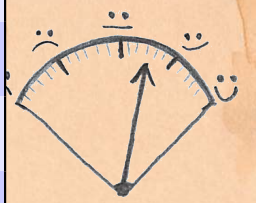
The UNIVERSITY of VERMONT

Most frequent ANEW words:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

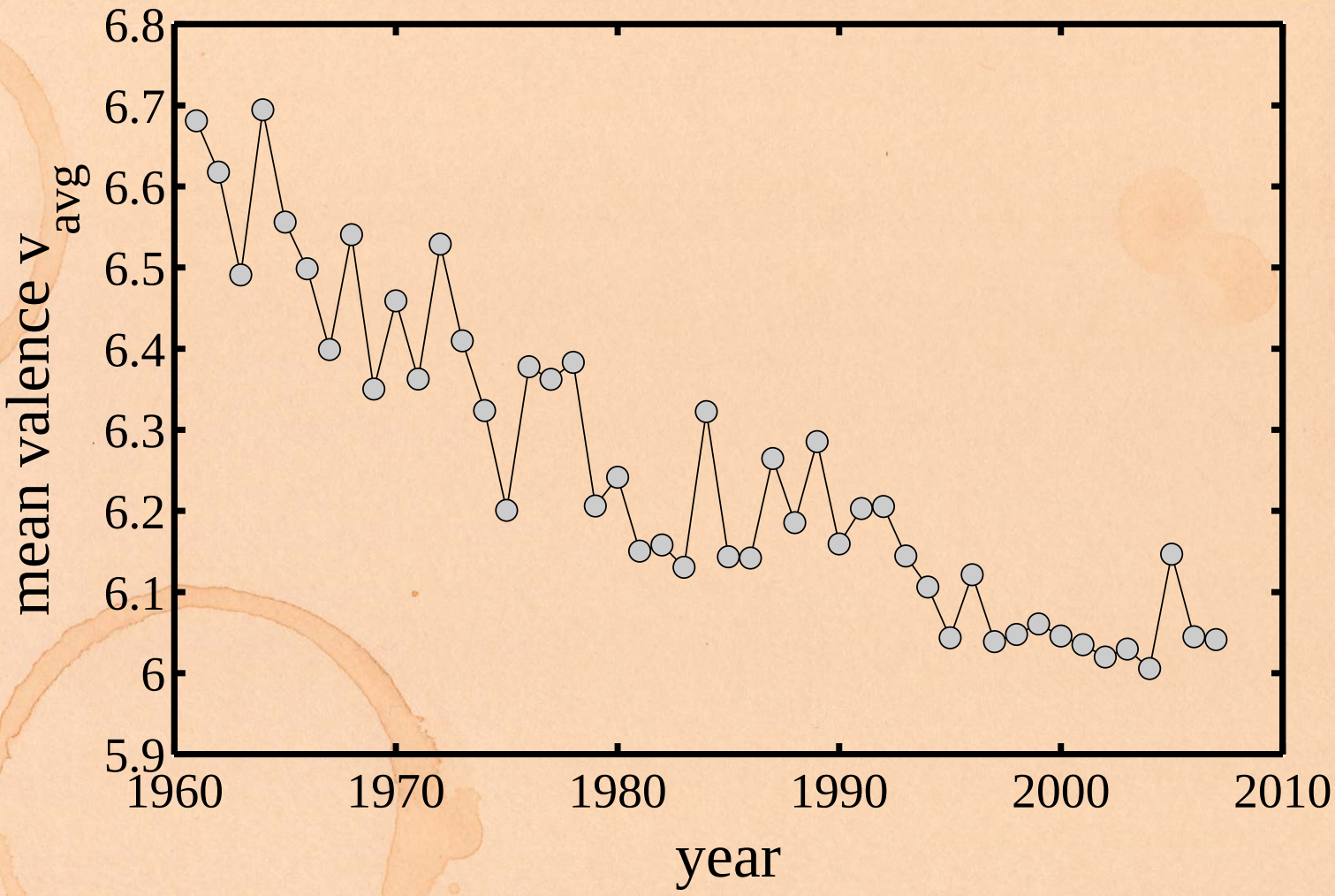
Rank	Song lyrics	Song titles
1	love (7.37%)	love (7.39%)
2	time (4.18%)	time (4.19%)
3	baby (2.75%)	baby (2.75%)
4	life (2.59%)	life (2.60%)
5	heart (2.14%)	heart (2.15%)

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%)

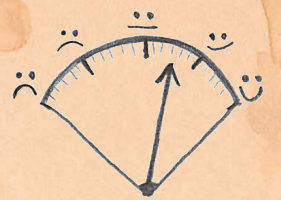


Text:	h_{avg}	Words with a similar score:
Soul/Gospel lyrics [7]	6.9	chocolate (6.88), leisurely (6.88), penthouse (6.81)
Pop lyrics [7]	6.7	dream (6.73), honey (6.73), sugar (6.74)
Dante's Paradise [1]	6.5	muffin (6.57), rabbit (6.57), smooth (6.58)
Tweets, 9/9/2008 to 12/31/2010	6.4	thought (6.39), face (6.39), blond (6.42)
Rock lyrics [7]	6.3	church (6.28), tree (6.32), air (6.34)
Enron Emails [2]	6.2	clouds (6.18), alert (6.20), computer (6.24)
State of the Union Messages [7]	6.1	grass (6.12), idol (6.12), bottle (6.15)
New York Times (1987–2007) [14]	6.0	hotel (6.00), tennis (6.02), wonder (6.03)
Blogs [7]	5.8	owl (5.80), whistle (5.81), humble (5.86)
Dante's Inferno [1]	5.5	glacier (5.50), repentant (5.53), mischief (5.57)
Heavy Metal lyrics [7]	5.4	lamp (5.41), elevator (5.44), truck (5.47)

Song Lyrics—average happiness (valence)

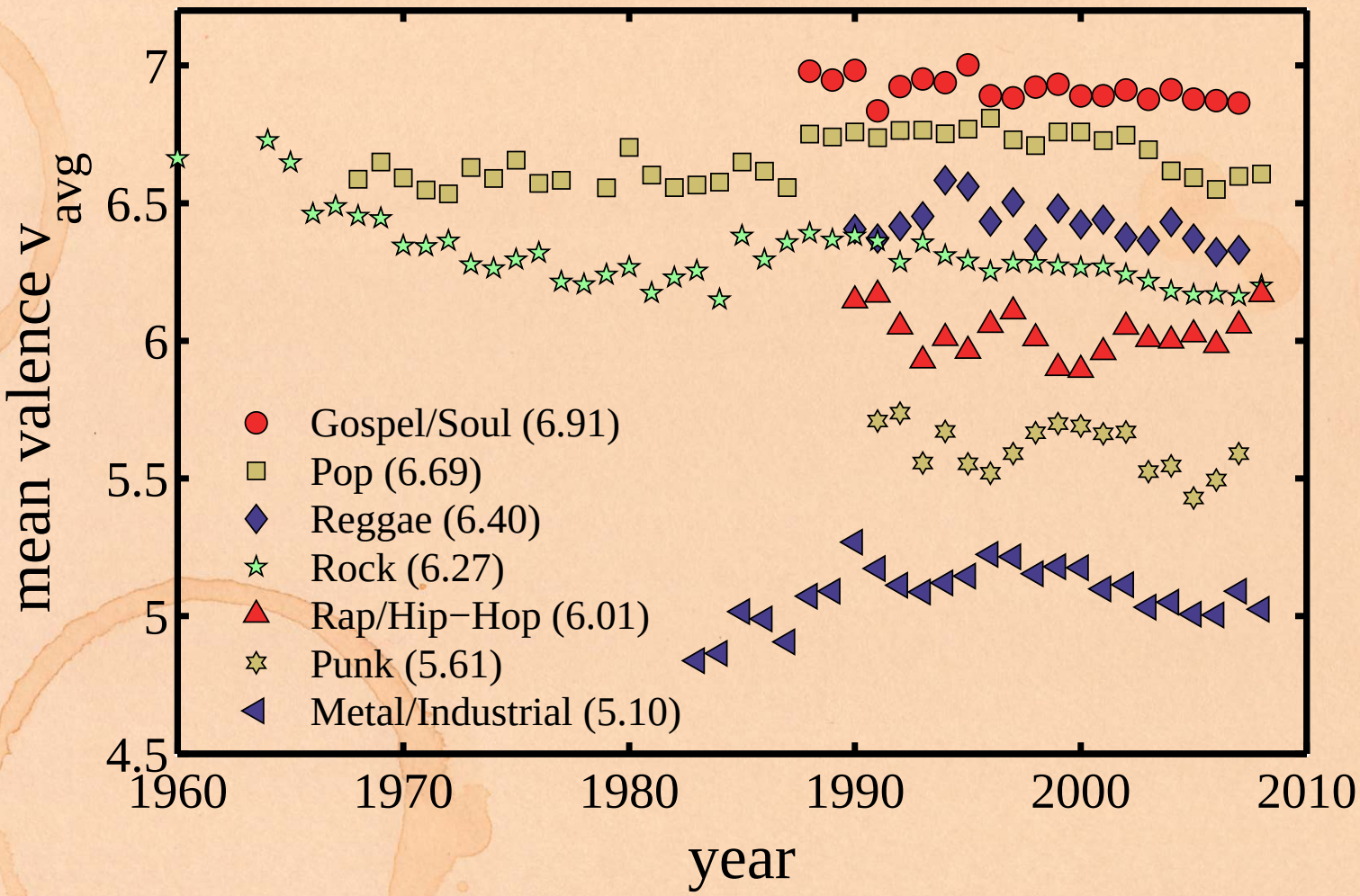


Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

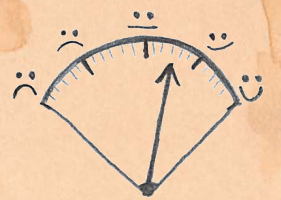


The UNIVERSITY of VERMONT

Song Lyrics—average valence of genres:



Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

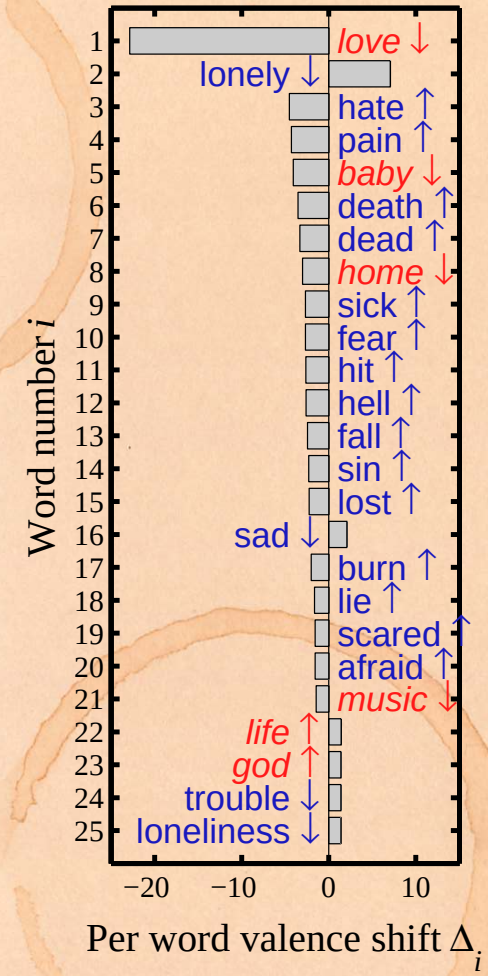


The UNIVERSITY of VERMONT

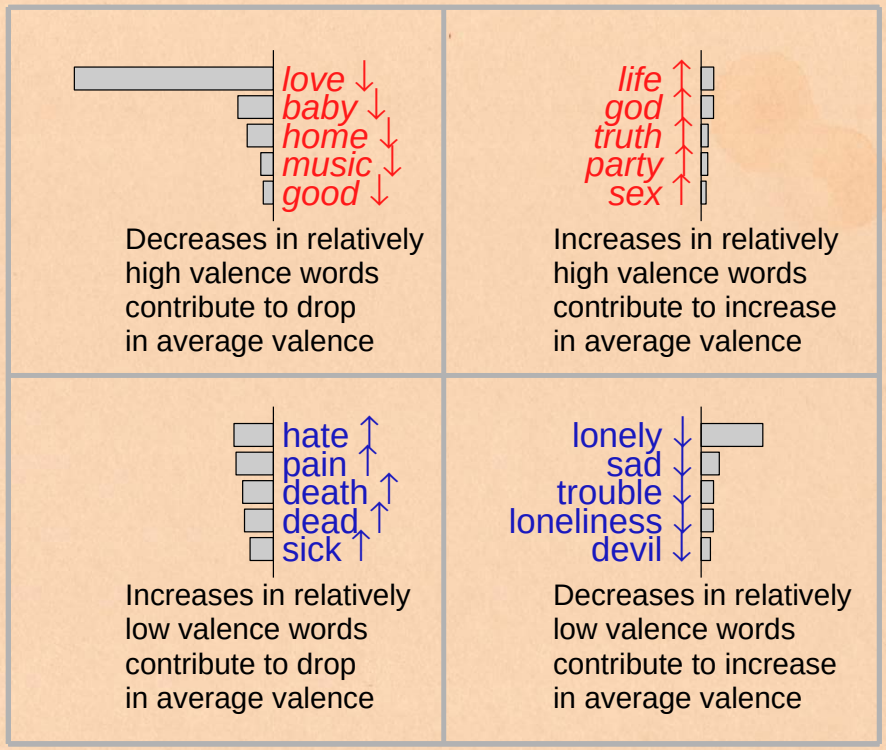
Happiness Word Shift Graph:

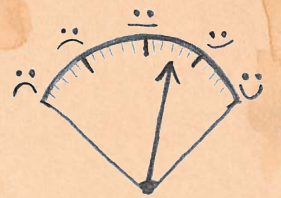
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

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



Key:



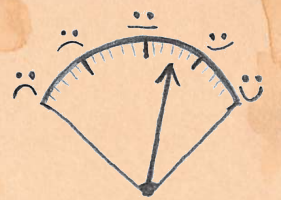


Top 16 of $\simeq$ 20,000 artists:

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)

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

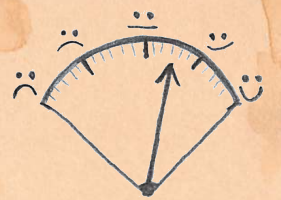


Bottom 16 of $\simeq$ 20,000 artists:

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

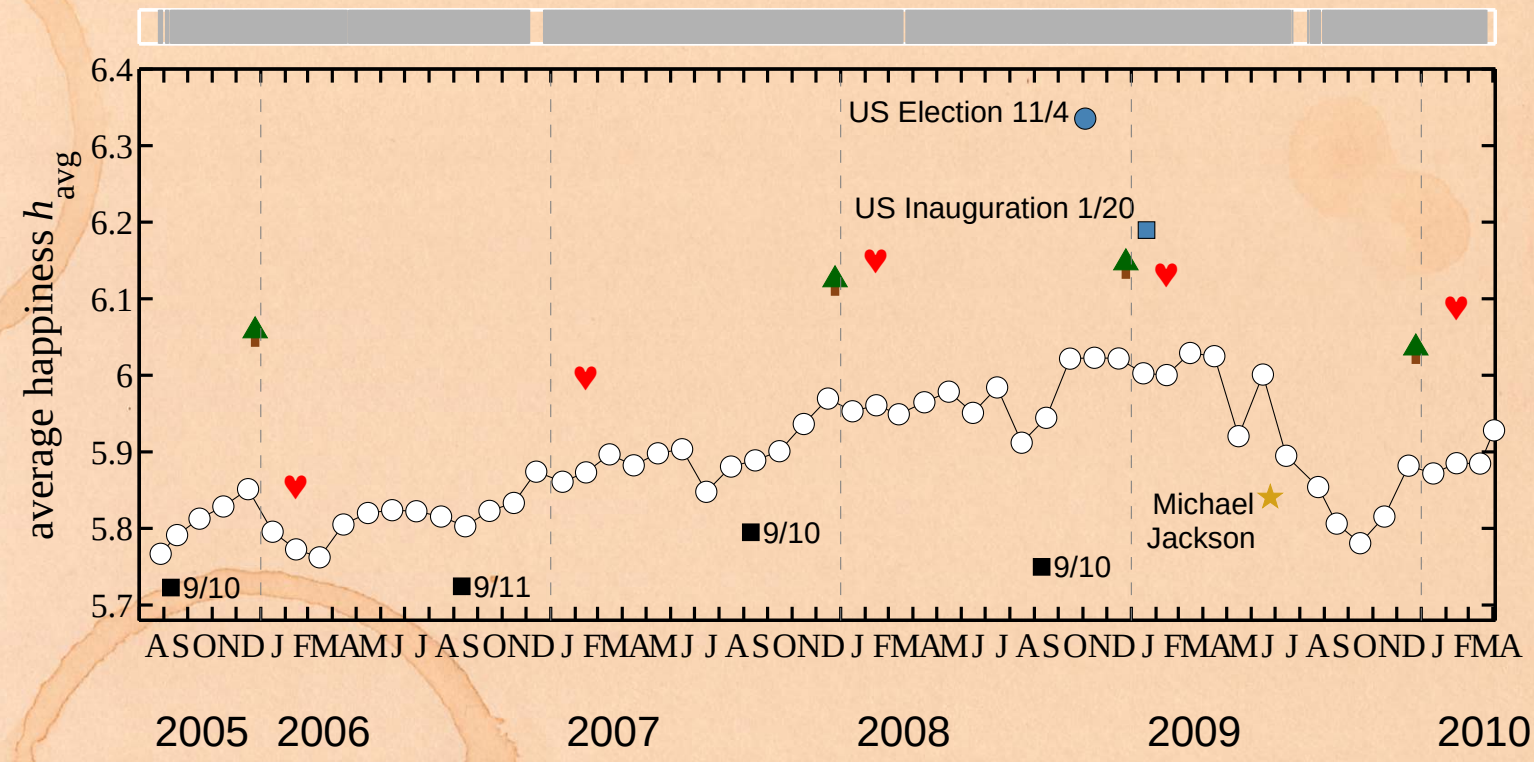
(criteria: $\geq$ 50 songs and $\geq$ 1000 ANEW words)

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



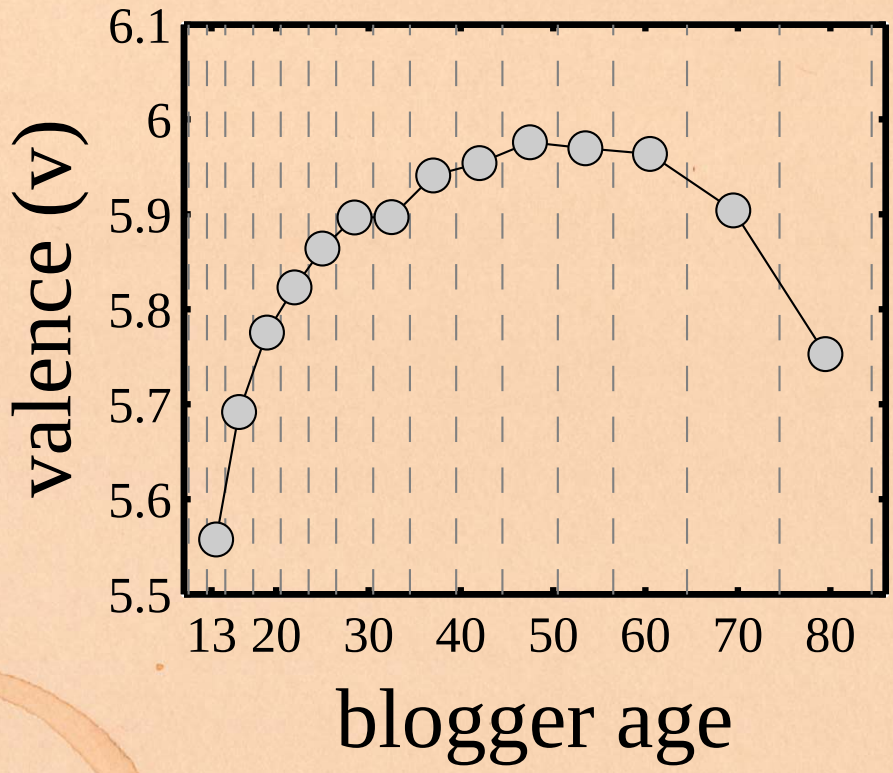
Blogs—Overall trend

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

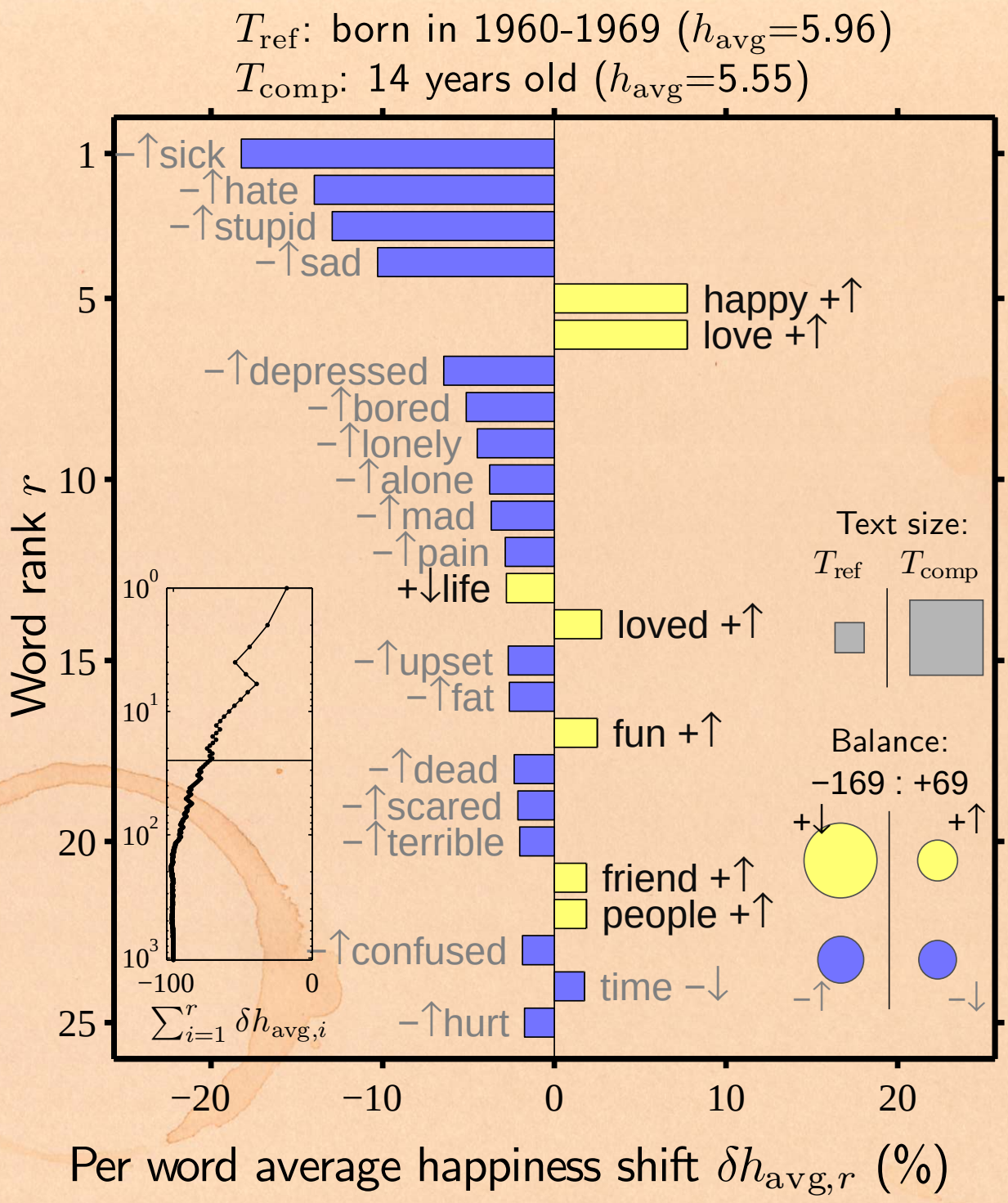


Blogs

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

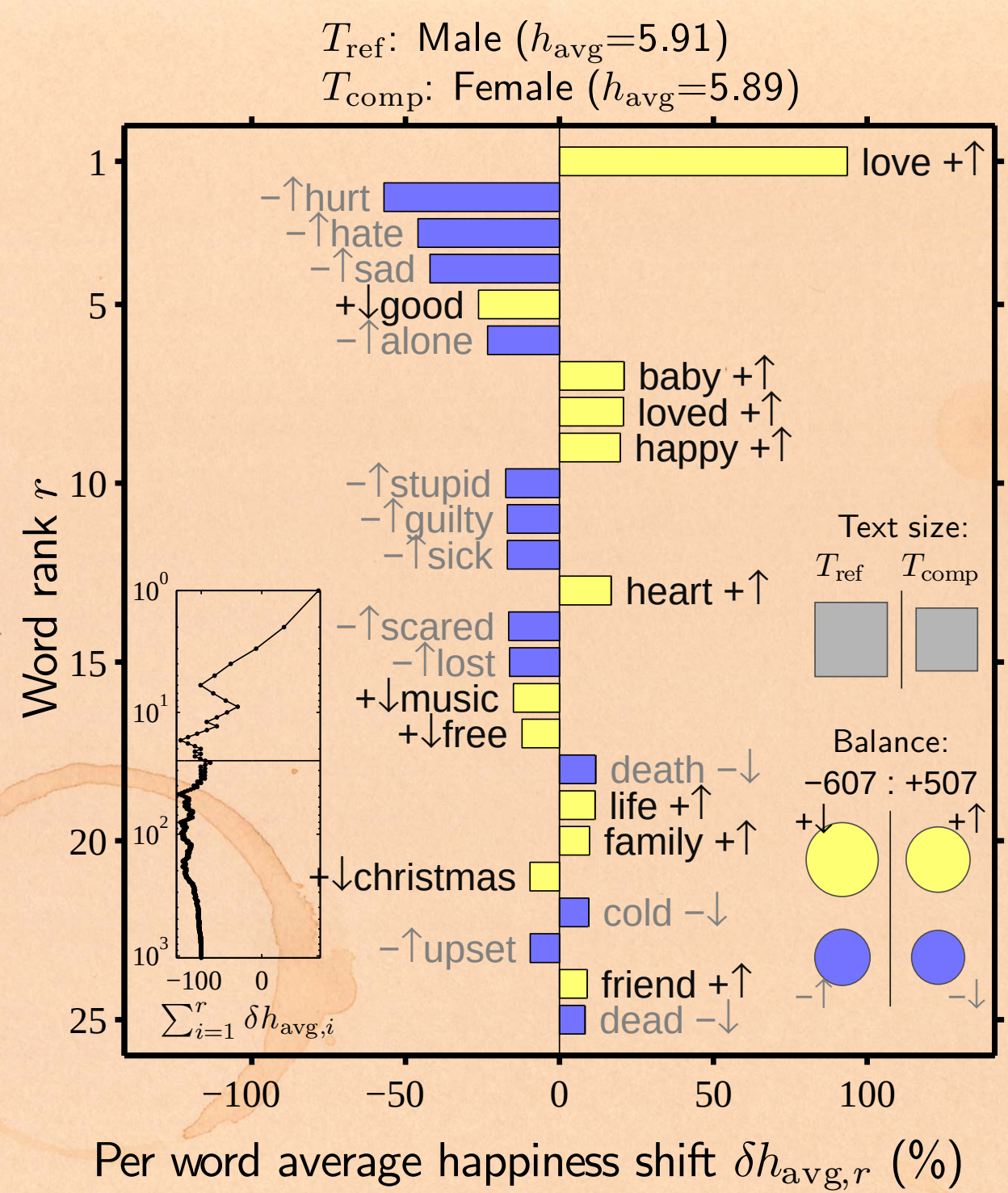


- Average happiness as a function of the age bloggers report they will turn in the year of their posting.



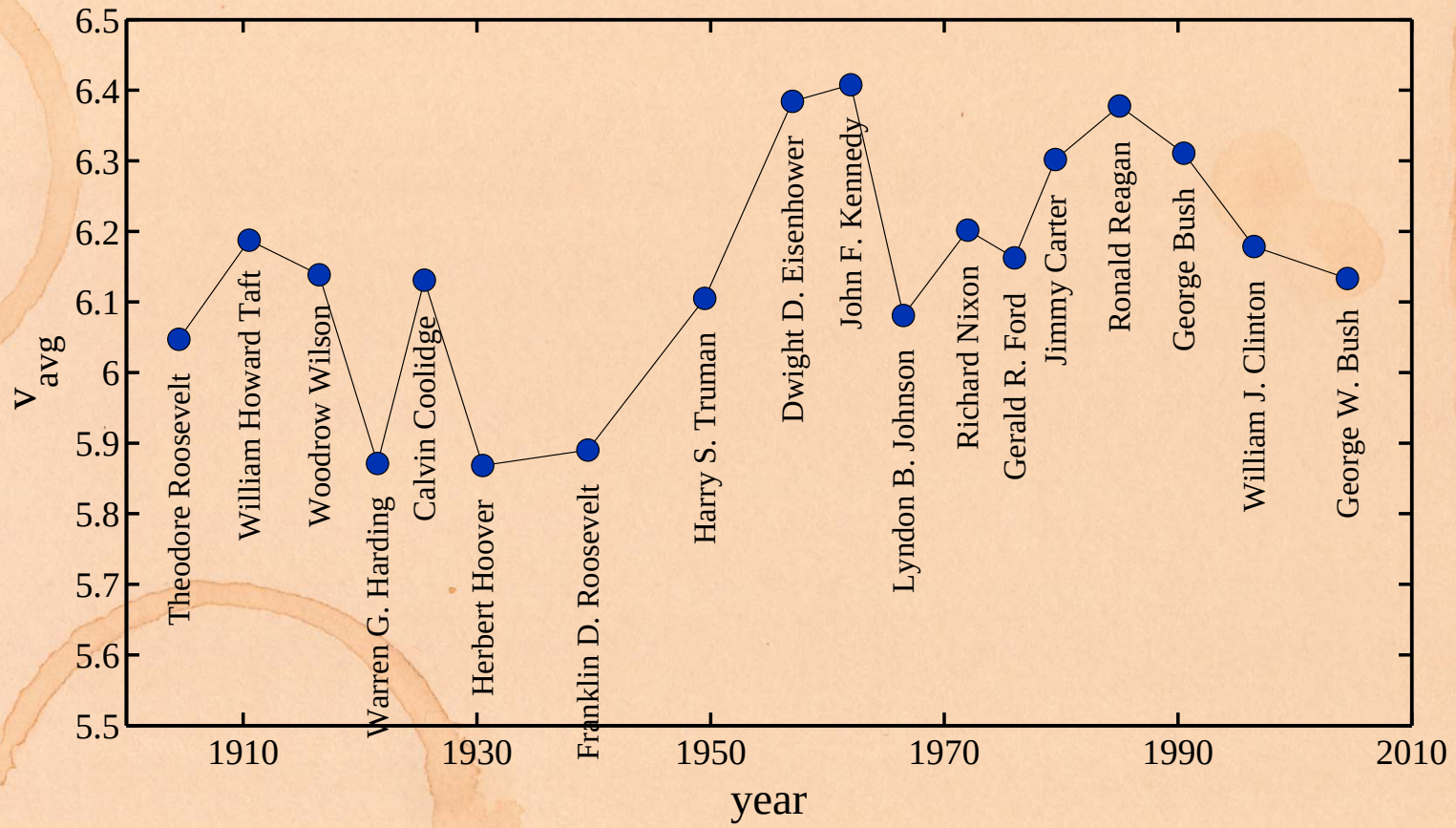
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



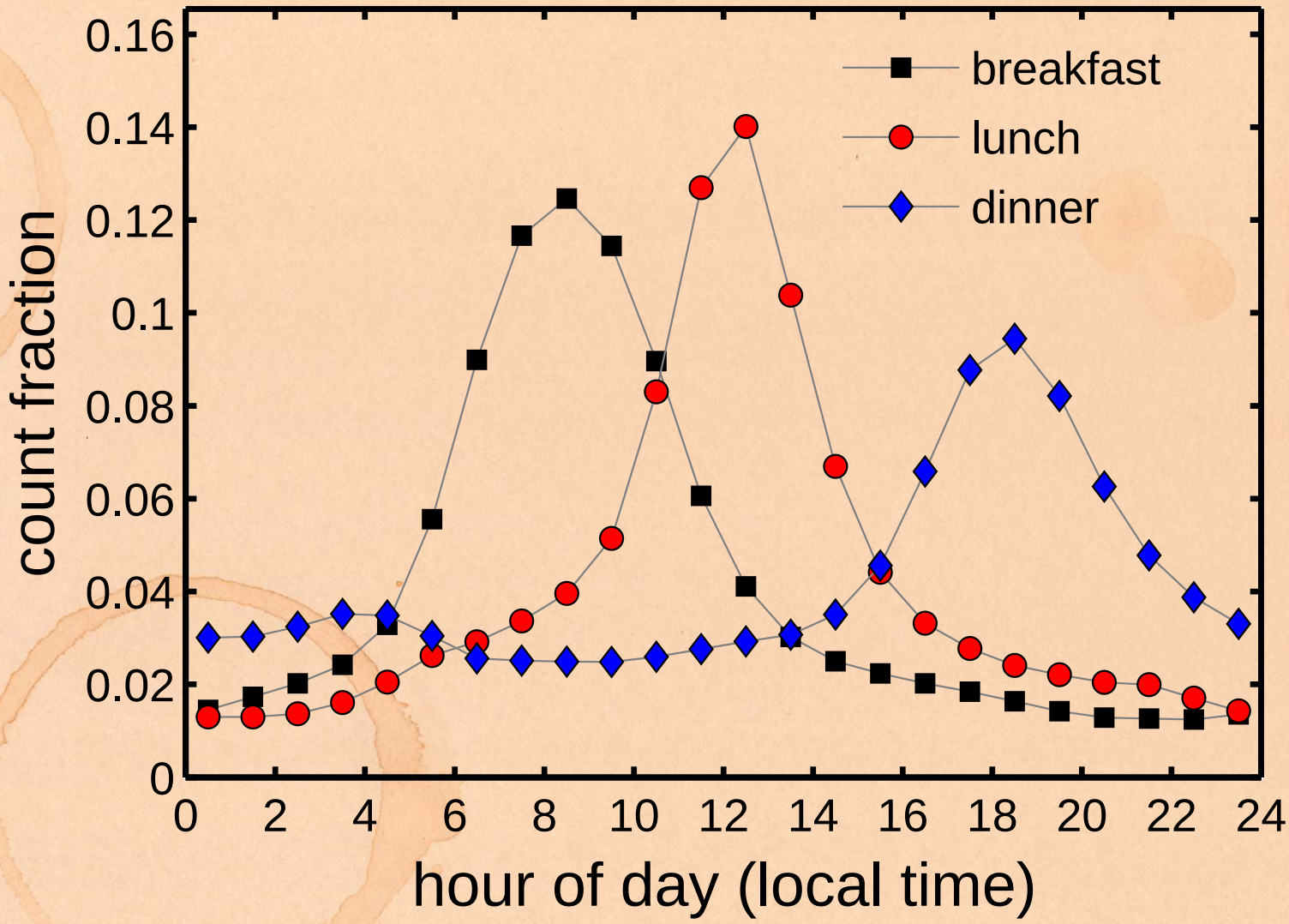
Presidential happiness:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



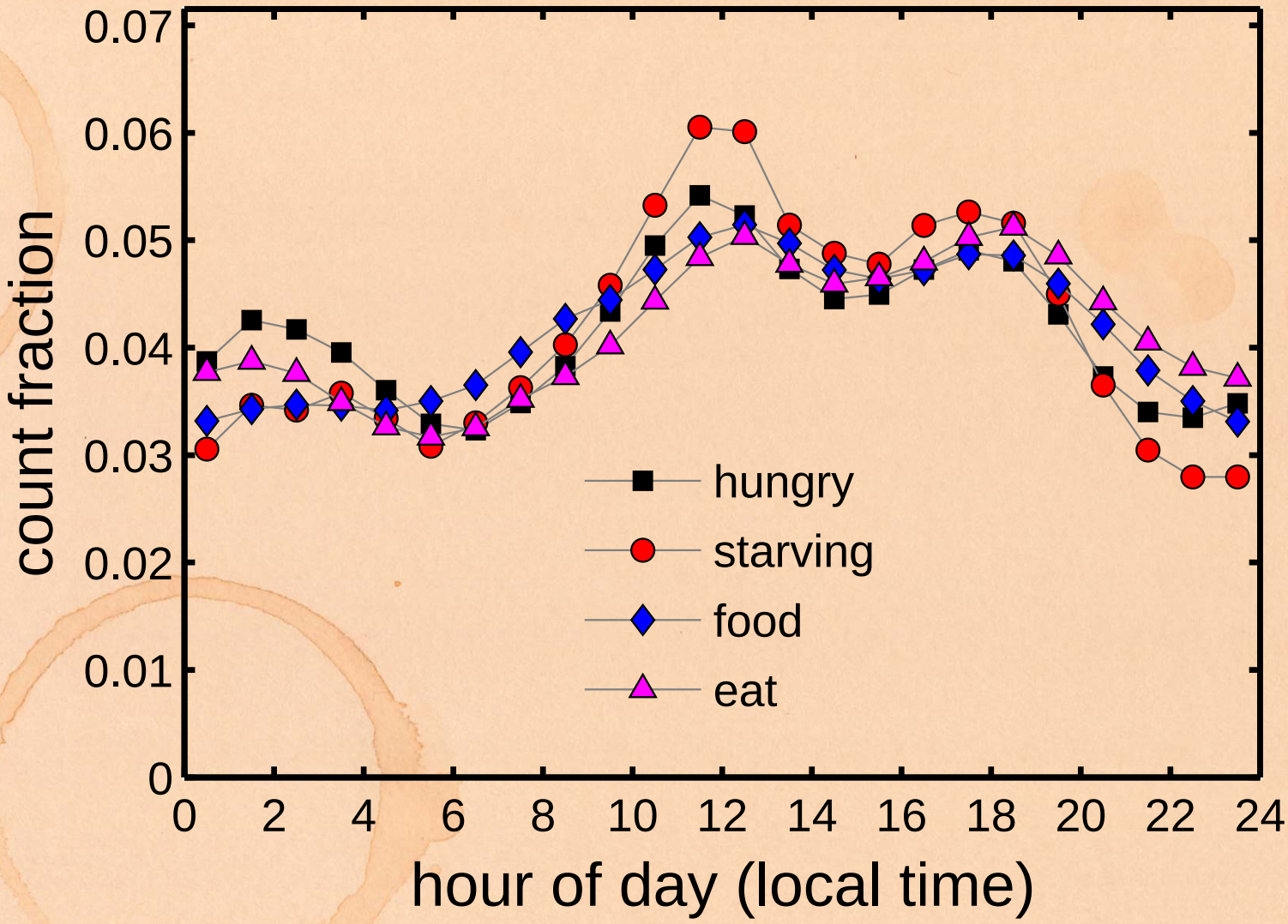
Twitter—living in the now:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



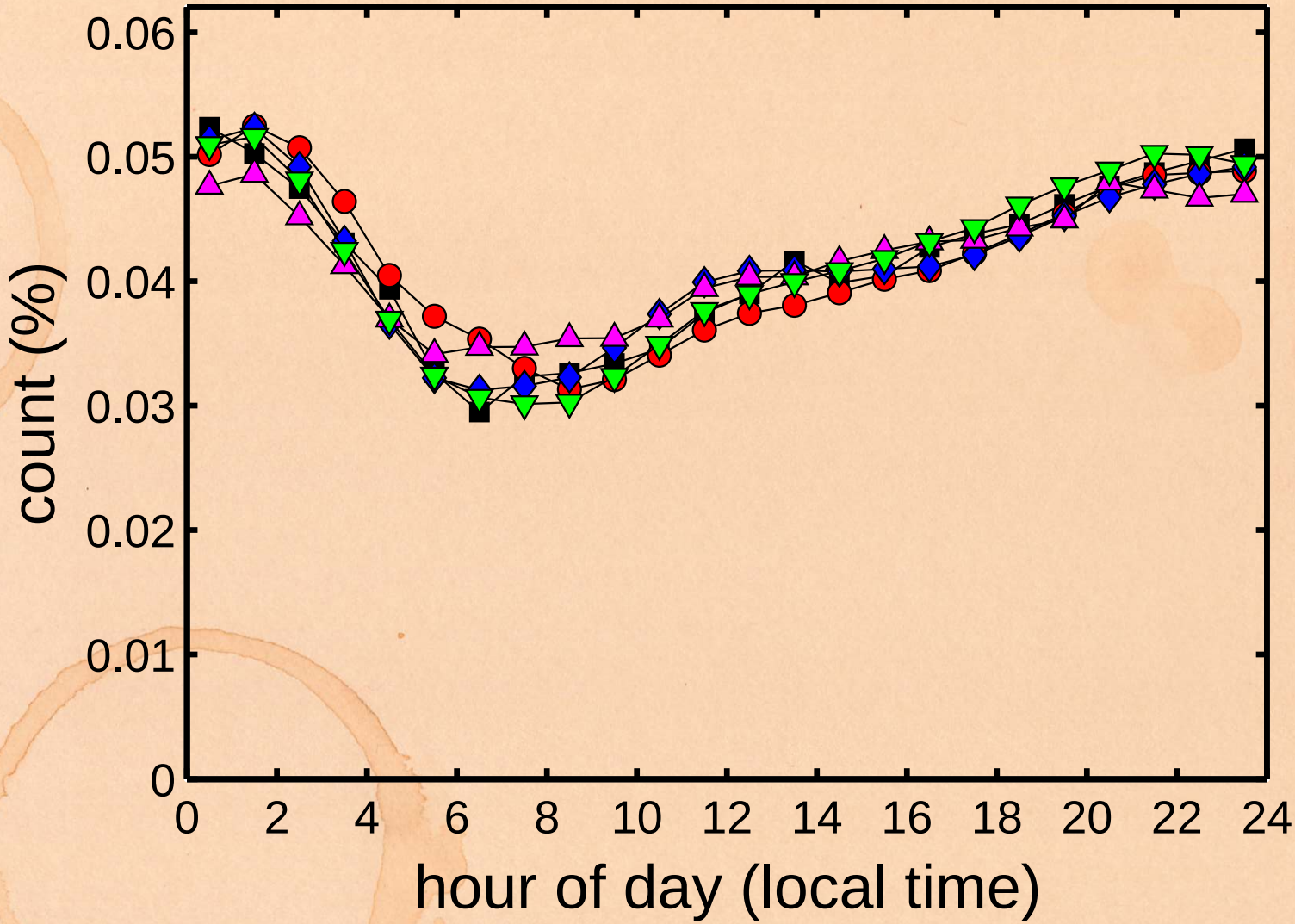
Twitter—living in the now:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



Twitter—living in the now:

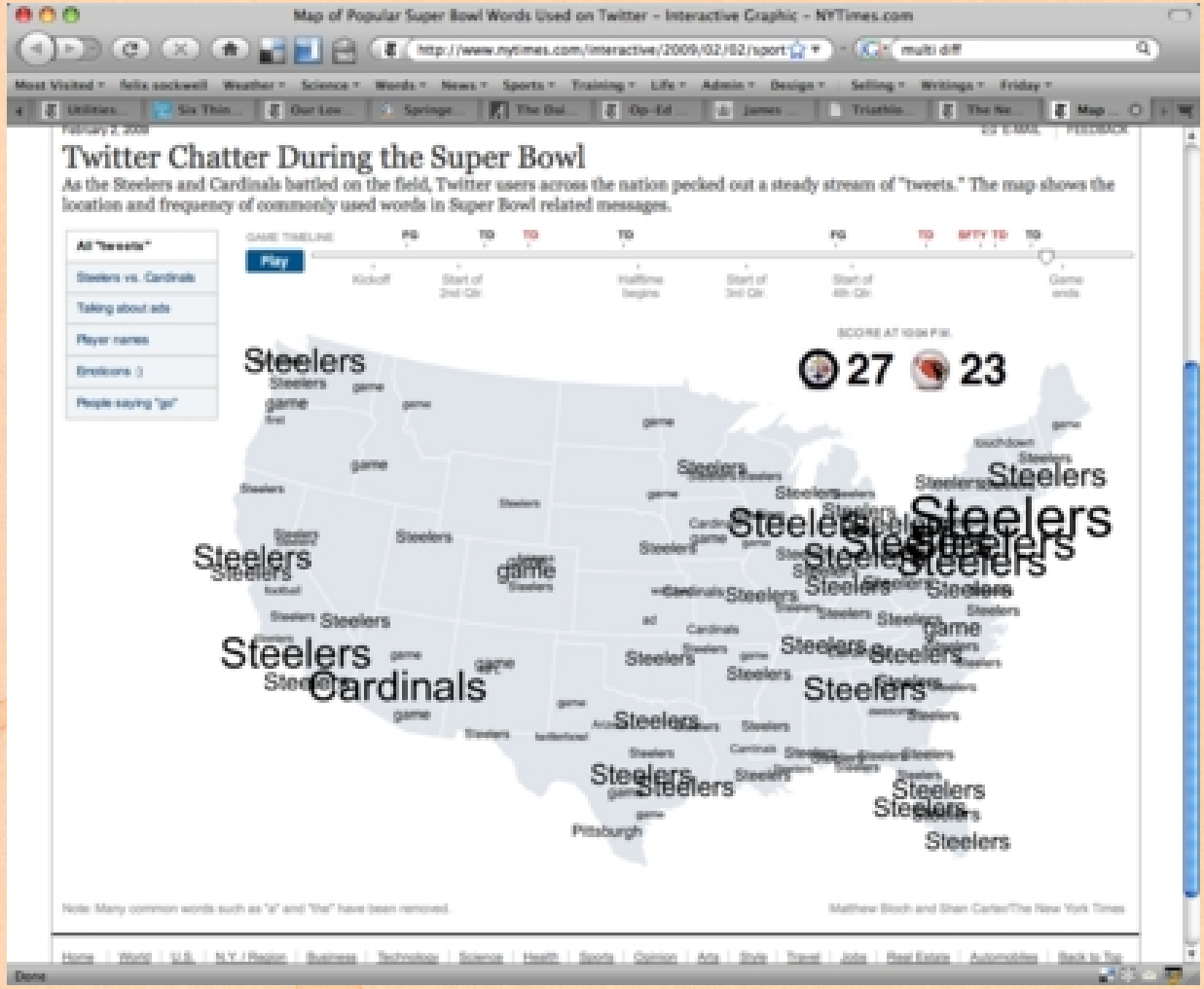
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



A few words you can't say on television.

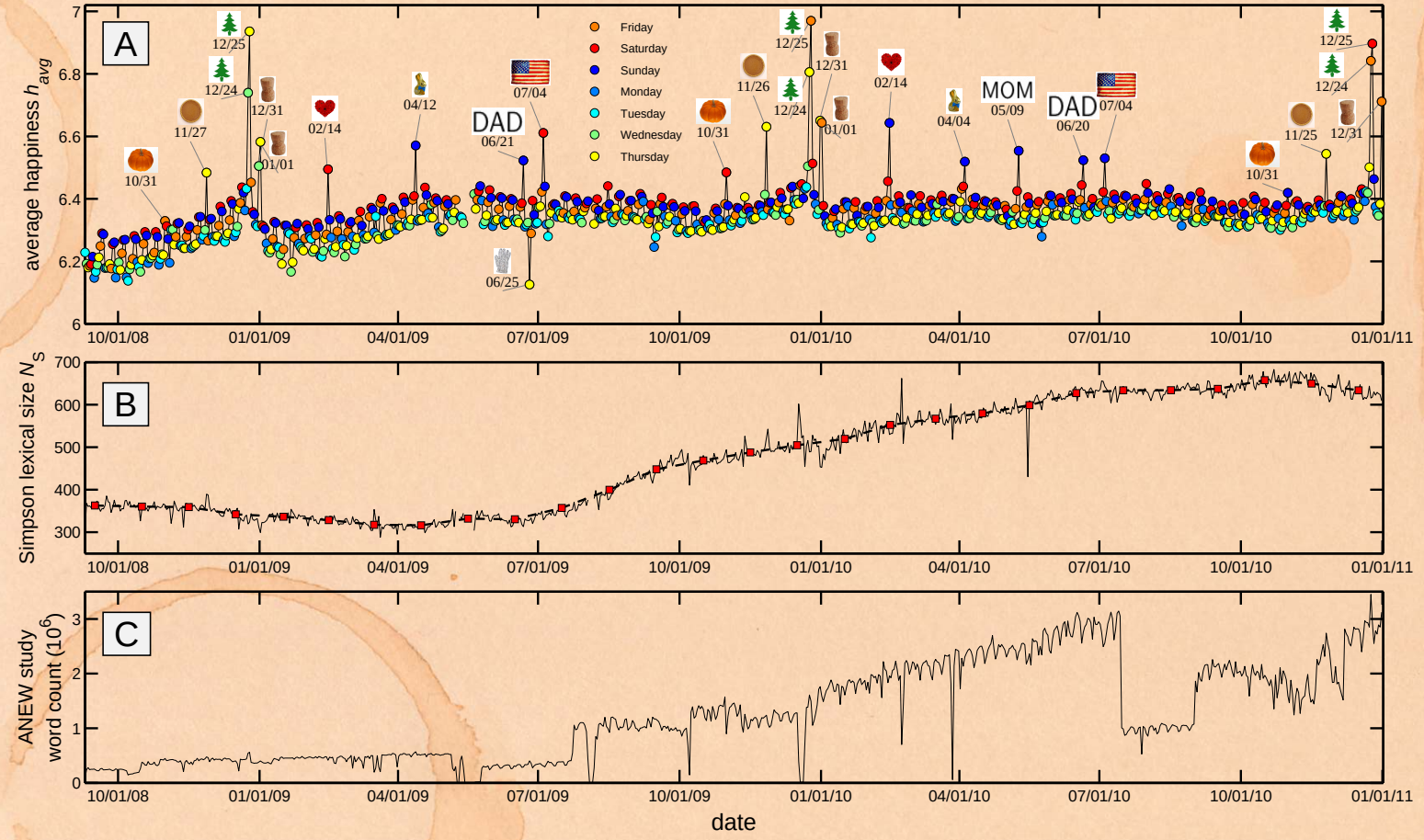
Twitter—living in the now:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



Tweeting the Superbowl (⊞) [NY Times]

Twitter—overall time series:

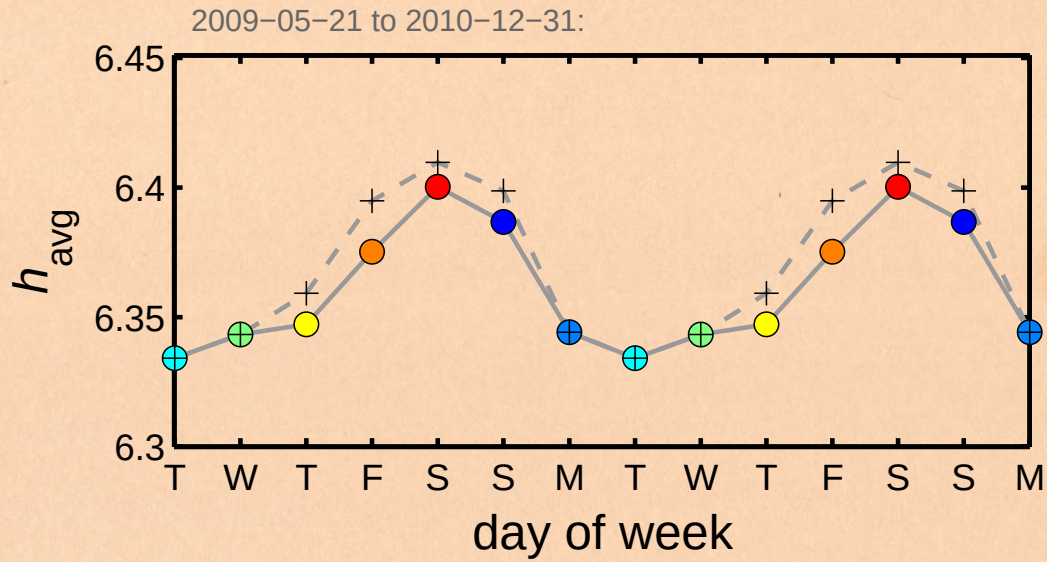


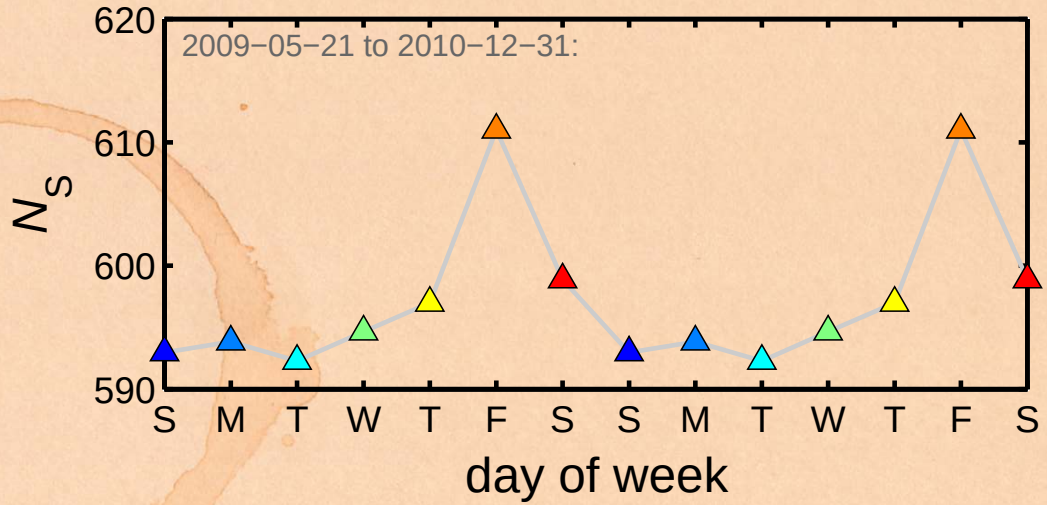
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

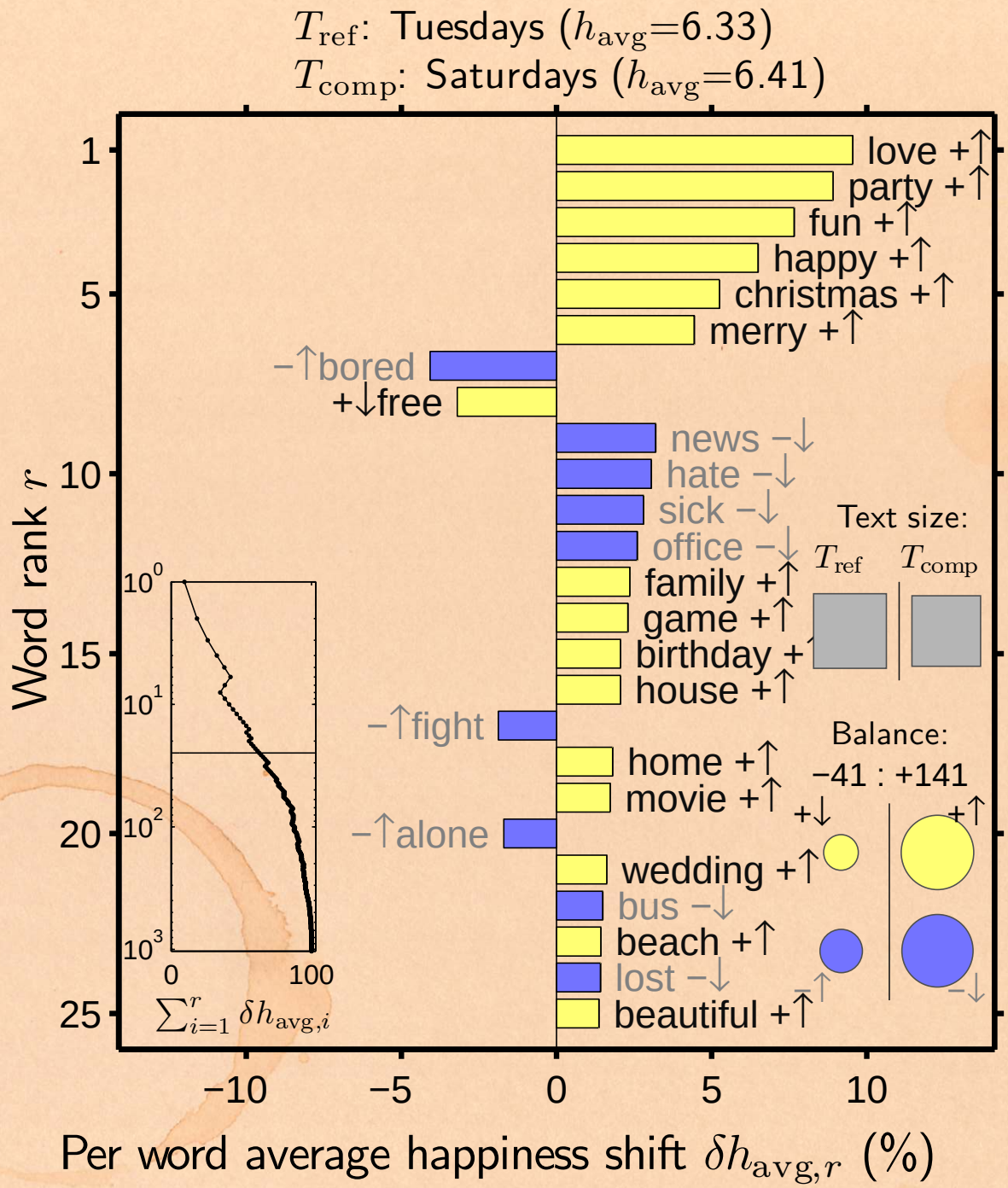
The UNIVERSITY of VERMONT

Twitter—weekly time series:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References





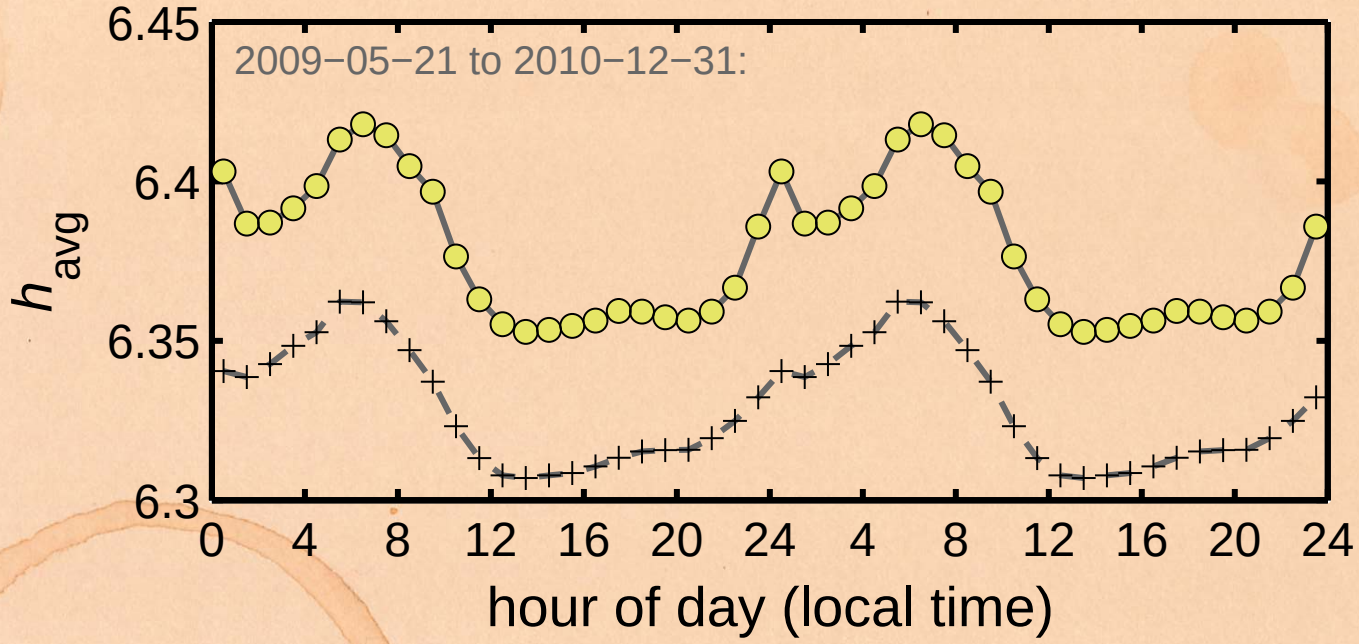


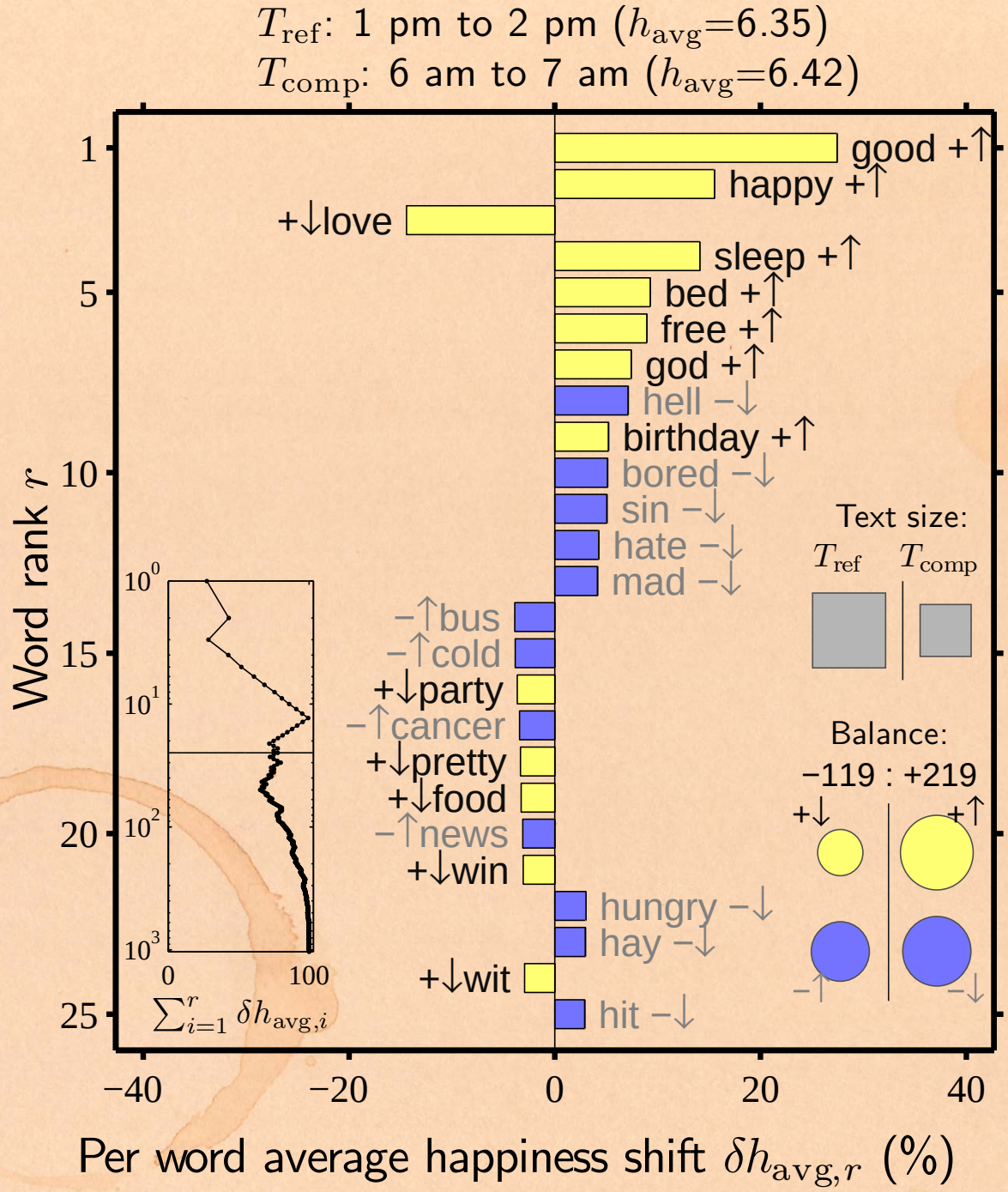
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The UNIVERSITY of VERMONT

Twitter—daily time series:

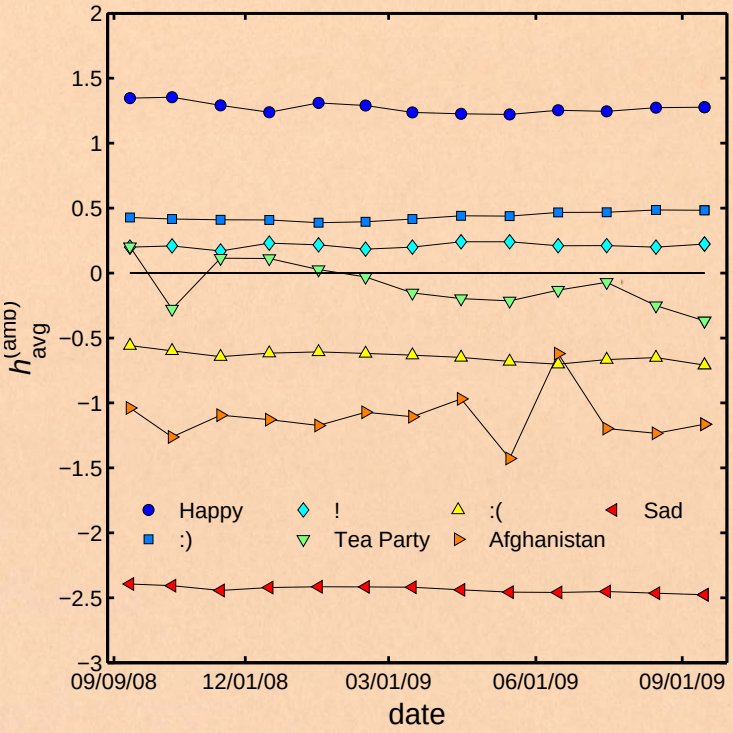
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

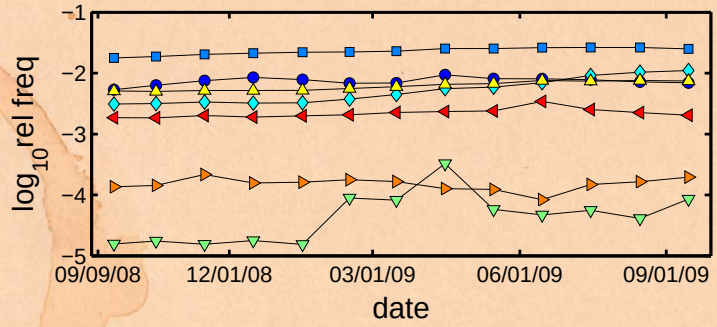




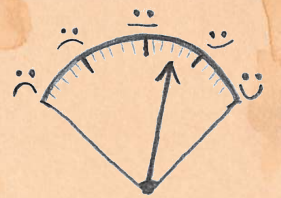
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The UNIVERSITY of VERMONT



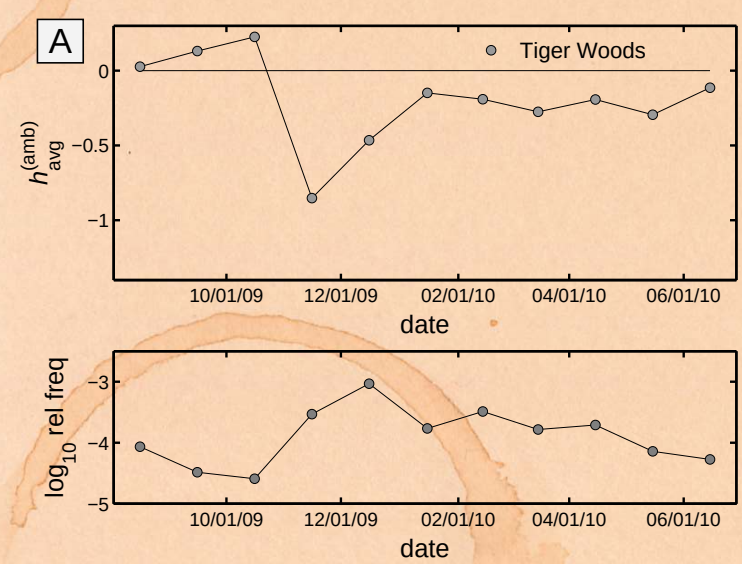


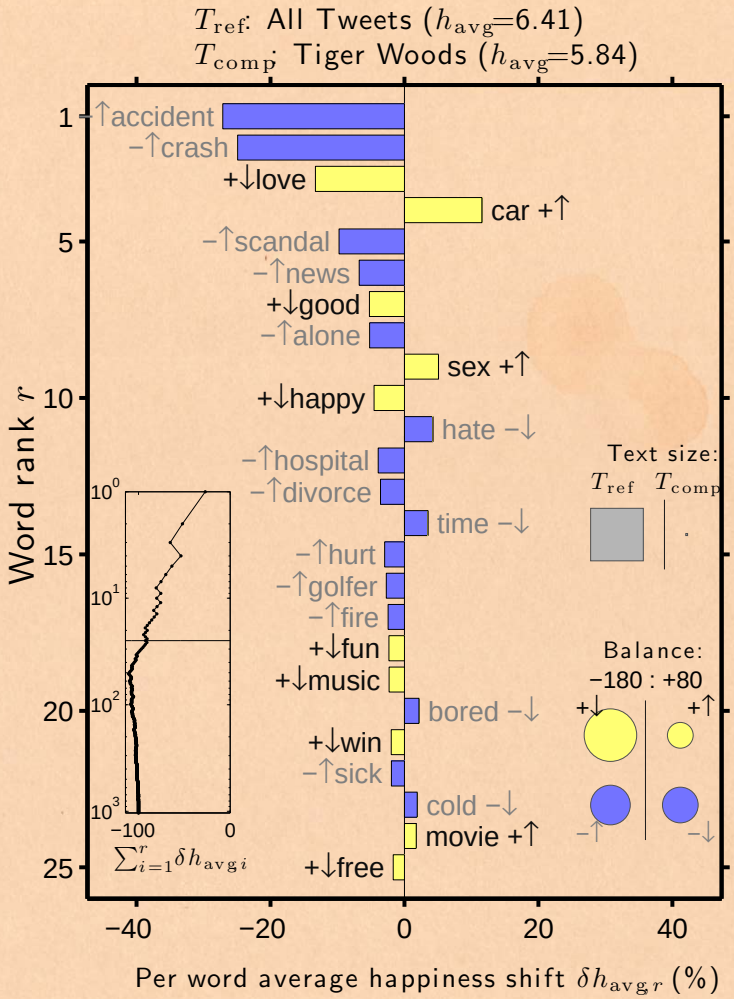
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

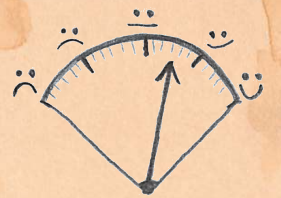


The UNIVERSITY of VERMONT

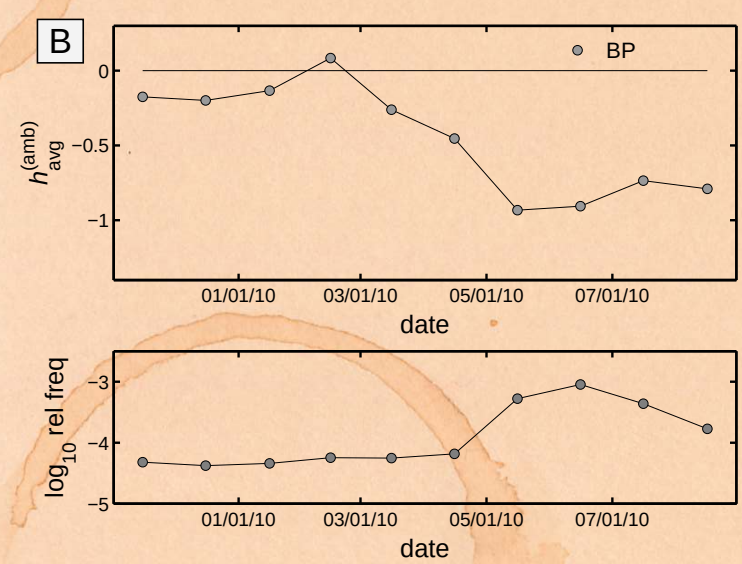
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

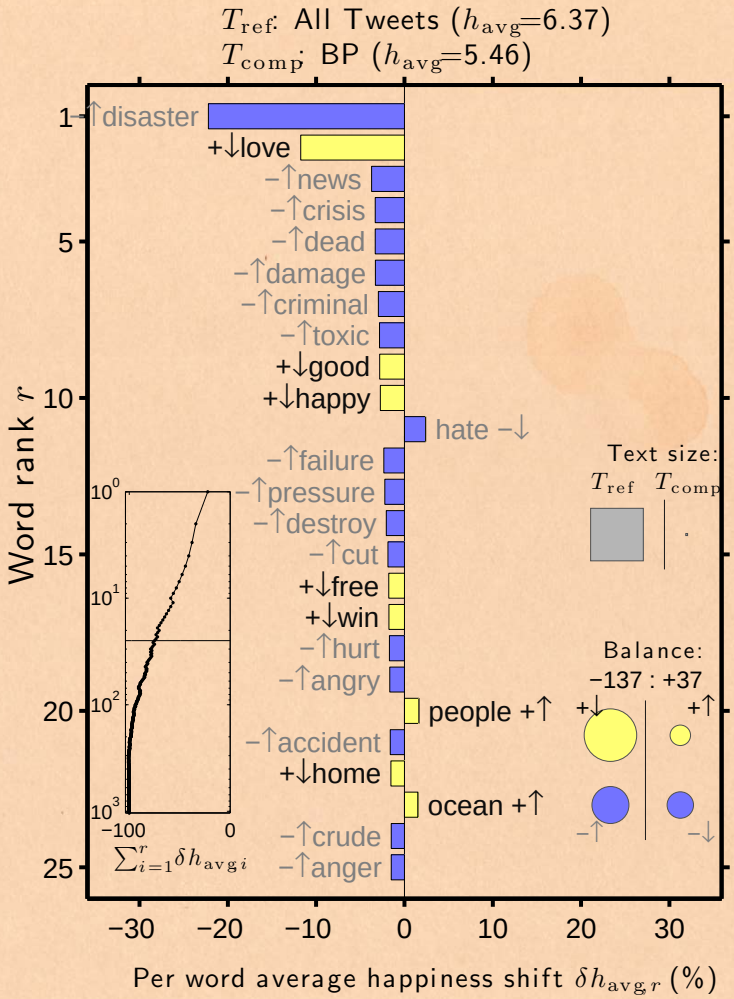


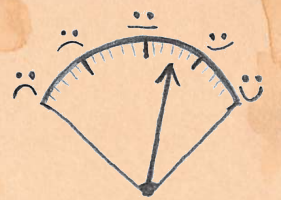




Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References





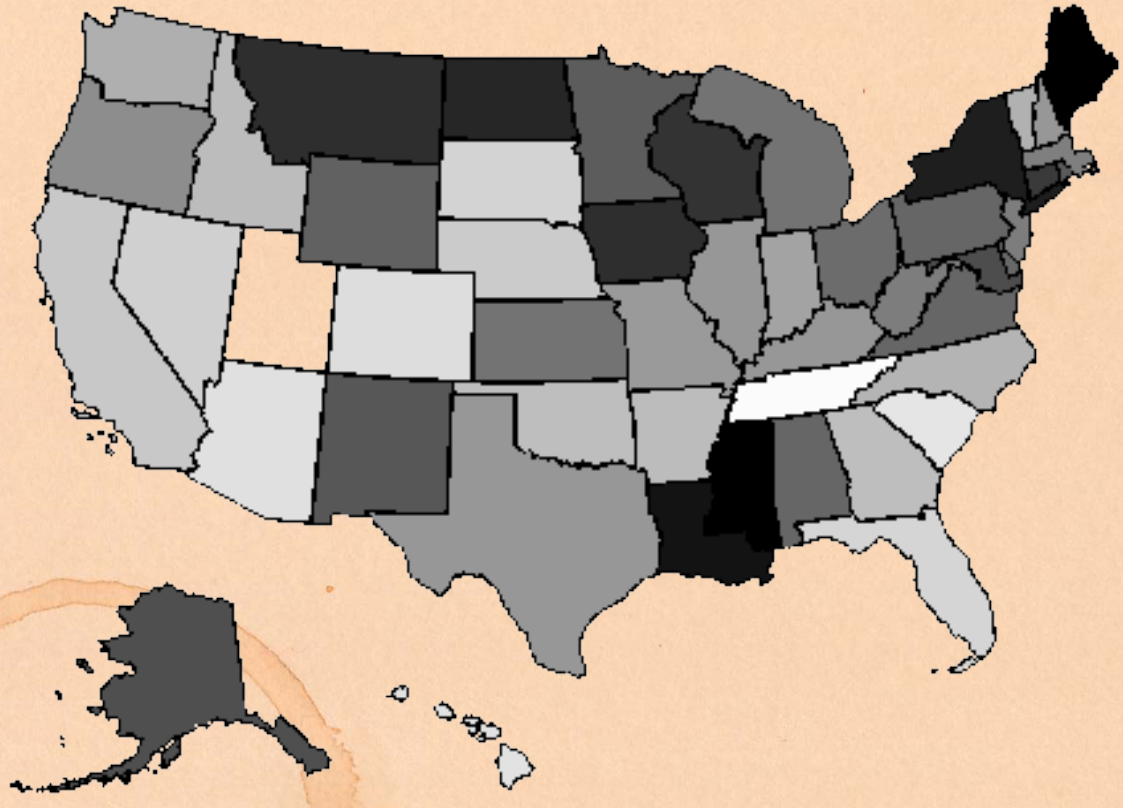


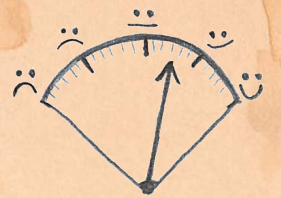
Word	$h_{\text{avg}}^{(\text{amb})}$	Total Tweets	Total ANEW
1. love	+1.42	46,687,476 (6)	85,269,499 (5)
2. happy	+1.32	16,541,968 (13)	32,442,529 (8)
3. win	+1.26	7,981,856 (26)	14,640,728 (20)
4. kiss	+1.21	1,697,405 (59)	3,162,330 (48)
5. cash	+1.21	1,279,236 (63)	2,468,496 (51)
6. vacation	+1.11	934,501 (67)	1,783,270 (56)
7. Christmas	+1.03	4,887,968 (35)	10,645,630 (25)
8. God	+0.95	8,576,364 (25)	17,867,768 (16)
9. party	+0.93	6,438,886 (29)	12,090,597 (23)
10. sex	+0.89	3,551,767 (39)	7,087,972 (31)
11. Valentine	+0.85	247,288 (84)	464,914 (75)
12. family	+0.79	5,014,816 (32)	10,629,361 (26)
13. sun	+0.65	2,385,348 (52)	4,602,627 (44)
14. life	+0.50	14,006,454 (17)	27,770,768 (10)
15. hope	+0.48	11,833,337 (18)	22,952,366 (13)
16. heaven	+0.43	741,878 (71)	1,485,702 (59)
17. :)	+0.42	10,470,483 (20)	6,787,678 (35)
18. income	+0.36	510,425 (76)	418,161 (77)
19. friends	+0.33	7,669,719 (27)	7,541,106 (29)
20. snow	+0.32	2,596,165 (49)	5,011,785 (40)
21. :-)	+0.32	1,680,165 (60)	1,102,512 (67)
22. night	+0.29	17,089,505 (12)	17,606,796 (17)
23. vegan	+0.28	183,889 (90)	178,676 (86)
24. Jesus	+0.27	2,027,720 (56)	1,673,992 (58)
25. girl	+0.25	10,070,132 (22)	19,886,691 (14)
26. USA	+0.23	2,157,172 (54)	1,204,585 (65)
27. you	+0.22	173,276,993 (3)	145,464,084 (2)
28. our	+0.21	14,062,465 (16)	14,437,899 (21)
29. ;)	+0.20	2,618,940 (48)	1,475,221 (60)
30. health	+0.20	2,575,543 (50)	4,950,202 (41)
31. tomorrow	+0.20	10,379,637 (21)	8,899,406 (28)
32. !	+0.16	3,463,257 (40)	1,385,072 (62)
33. summer	+0.13	2,998,785 (43)	2,554,459 (50)
34. we	+0.13	39,132,934 (7)	34,513,587 (7)
35. today	+0.13	25,588,506 (9)	23,619,518 (12)
36. man	+0.12	15,856,341 (14)	29,558,118 (9)
37. woman	+0.10	2,543,036 (51)	5,603,347 (39)
38. Stephen Colbert	+0.10	23,778 (99)	14,697 (99)
39. ;-)	+0.10	943,413 (66)	516,171 (73)
40. RT	+0.06	339,055,724 (1)	142,219,359 (3)
41. coffee	+0.04	2,800,972 (46)	2,399,867 (52)
42. church	+0.03	1,812,251 (58)	3,452,171 (45)
43. work	+0.02	18,415,618 (11)	16,191,802 (18)
44. I	+0.02	307,960,343 (2)	282,865,043 (1)
45. yes	+0.02	11,593,356 (19)	7,499,840 (30)
46. them	0.00	15,352,295 (15)	14,398,889 (22)
47. hot	-0.01	7,122,144 (28)	6,286,163 (37)
48. boy	-0.01	4,933,333 (33)	9,670,512 (27)
49. yesterday	-0.01	3,077,761 (42)	2,852,623 (49)
50. Michael Jackson	-0.02	825,979 (70)	571,442 (71)

Word	$h_{\text{avg}}^{(\text{amb})}$	Total Tweets	Total ANEW
51. me	-0.06	144,342,098 (4)	88,088,051 (4)
52. ?	-0.07	2,333,283 (53)	674,679 (69)
53. commute	-0.09	90,126 (94)	90,092 (92)
54. gay	-0.09	2,727,309 (47)	1,697,177 (57)
55. right	-0.10	19,166,480 (10)	15,850,283 (19)
56. school	-0.11	9,264,217 (24)	6,924,193 (34)
57. Republican	-0.13	229,773 (86)	188,338 (85)
58. they	-0.16	27,442,360 (8)	27,150,189 (11)
59. winter	-0.19	1,255,945 (64)	1,217,225 (64)
60. lose	-0.19	2,056,468 (55)	2,091,540 (53)
61. Jon Stewart	-0.20	52,084 (97)	33,086 (96)
62. gas	-0.22	1,022,879 (65)	812,029 (68)
63. no	-0.22	95,129,093 (5)	38,894,616 (6)
64. Democrat	-0.23	93,193 (93)	75,450 (93)
65. left	-0.27	4,893,634 (34)	4,611,878 (43)
66. Senate	-0.29	447,732 (78)	316,835 (80)
67. election	-0.30	560,184 (75)	375,055 (78)
68. Sarah Palin	-0.34	225,577 (87)	150,096 (88)
69. Obama	-0.35	2,981,150 (44)	1,998,326 (54)
70. economy	-0.36	608,878 (73)	460,834 (76)
71. Congress	-0.36	391,510 (79)	279,695 (81)
72. drugs	-0.39	509,606 (77)	469,091 (74)
73. Muslim	-0.42	215,300 (88)	146,506 (89)
74. George Bush	-0.43	32,341 (98)	23,102 (98)
75. climate	-0.44	364,177 (80)	229,129 (83)
76. Pope	-0.51	152,320 (91)	135,955 (90)
77. oil	-0.53	1,377,355 (62)	1,148,990 (66)
78. I feel	-0.54	5,173,513 (31)	4,702,352 (42)
79. Glenn Beck	-0.54	113,991 (92)	101,090 (91)
80. Islam	-0.54	187,223 (89)	70,311 (94)
81. :-(	-0.65	341,141 (81)	244,215 (82)
82. :(	-0.70	2,907,145 (45)	1,891,225 (55)
83. flu	-0.75	901,403 (68)	639,000 (70)
84. rain	-0.78	3,233,464 (41)	5,959,903 (38)
85. BP	-0.78	582,167 (74)	326,100 (79)
86. mosque	-0.79	69,812 (95)	46,736 (95)
87. dark	-0.95	1,577,553 (61)	3,233,911 (47)
88. Lehman Brothers	-1.08	8,500 (100)	4,280 (100)
89. Goldman Sachs	-1.08	52,703 (96)	30,769 (97)
90. Afghanistan	-1.15	273,519 (83)	172,637 (87)
91. Iraq	-1.37	238,931 (85)	213,425 (84)
92. cold	-1.39	3,670,447 (36)	7,015,518 (32)
93. gun	-1.81	680,903 (72)	1,263,217 (63)
94. hate	-2.43	9,652,881 (23)	18,158,870 (15)
95. hell	-2.49	6,266,162 (30)	11,056,735 (24)
96. sick	-2.55	3,576,058 (37)	6,783,395 (36)
97. sad	-2.56	3,563,745 (38)	6,951,686 (33)
98. war	-2.63	1,955,901 (57)	3,417,588 (46)
99. depressed	-2.64	280,872 (82)	541,394 (72)
100. headache	-2.83	856,600 (69)	1,446,064 (61)

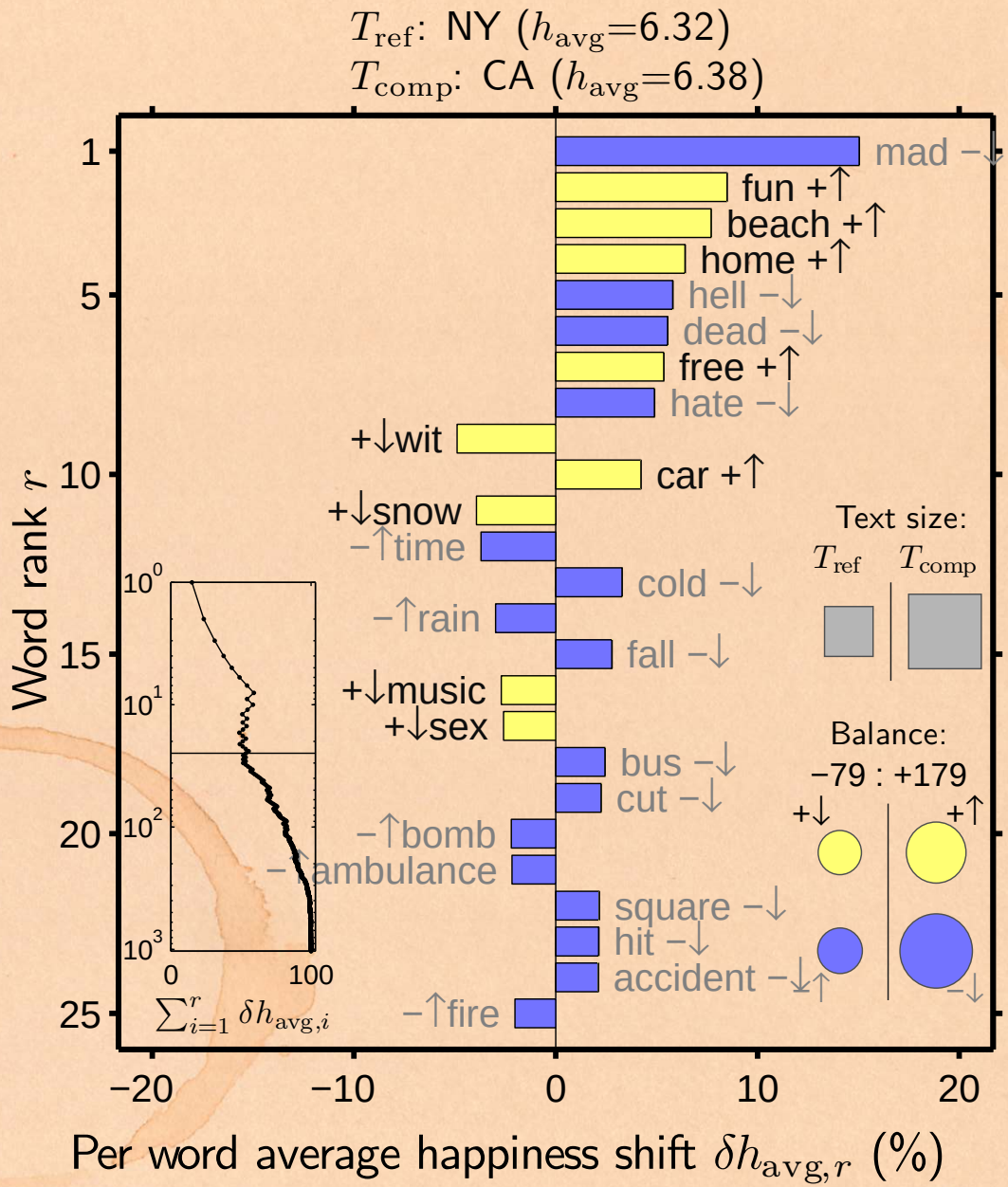
Twitter—location:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

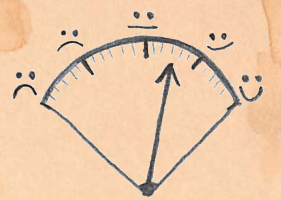




Twitter—location:

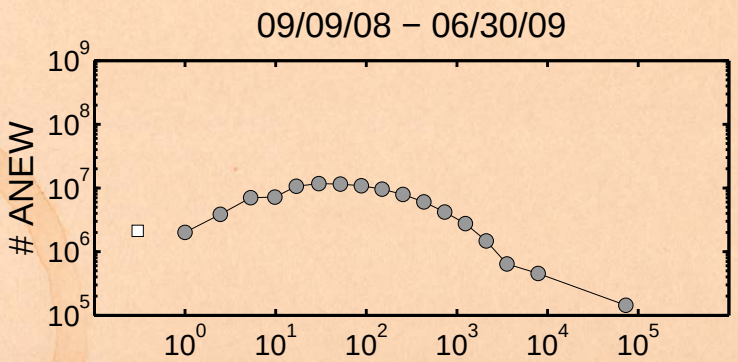


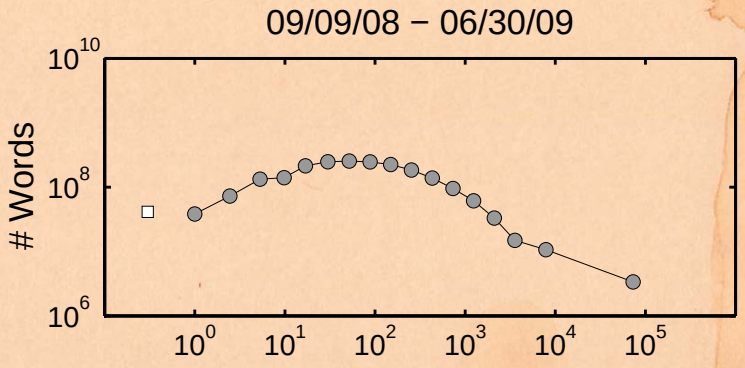
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

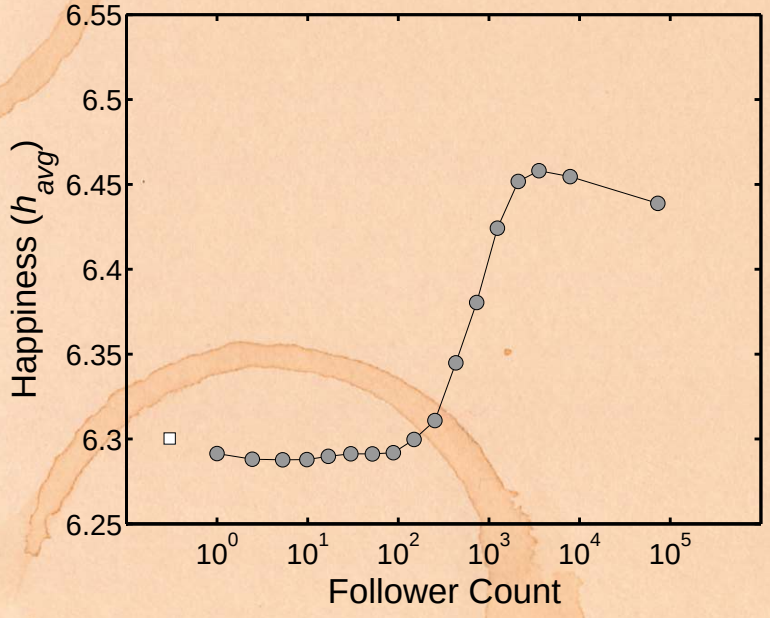


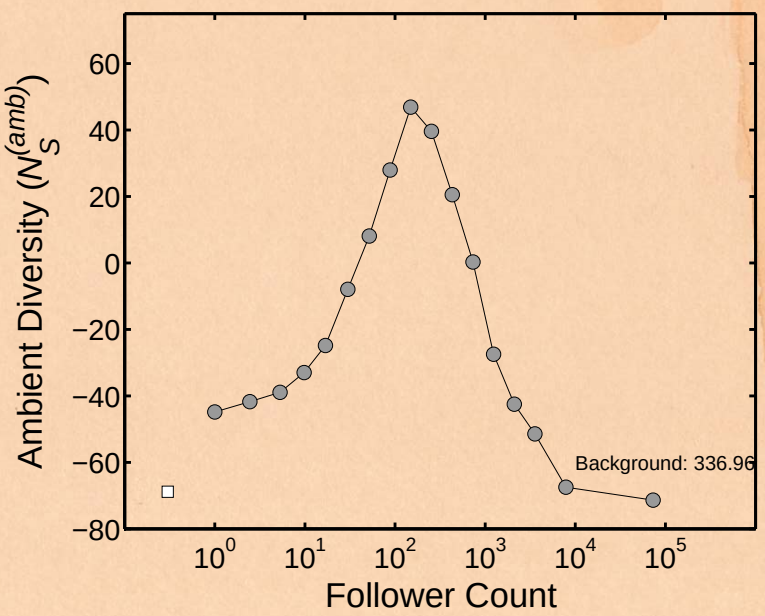
The UNIVERSITY of VERMONT

Twitter—popularity based on follower count:



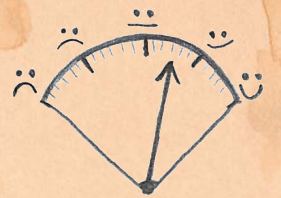






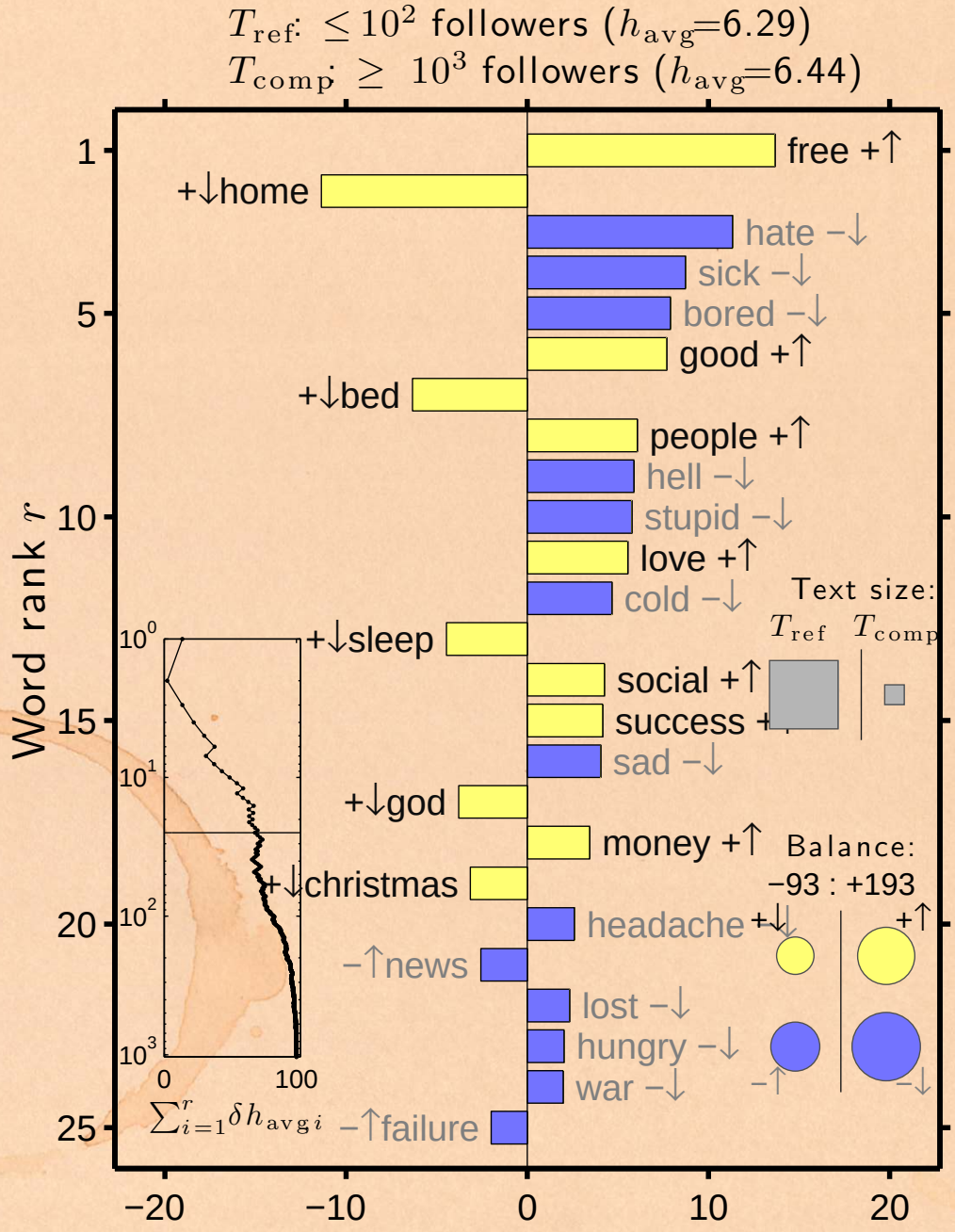
- Dunbar's number $\simeq$ 150.

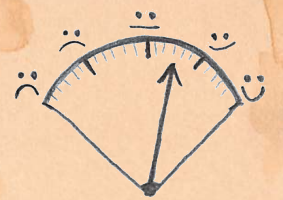
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



Twitter—popularity based on follower count:

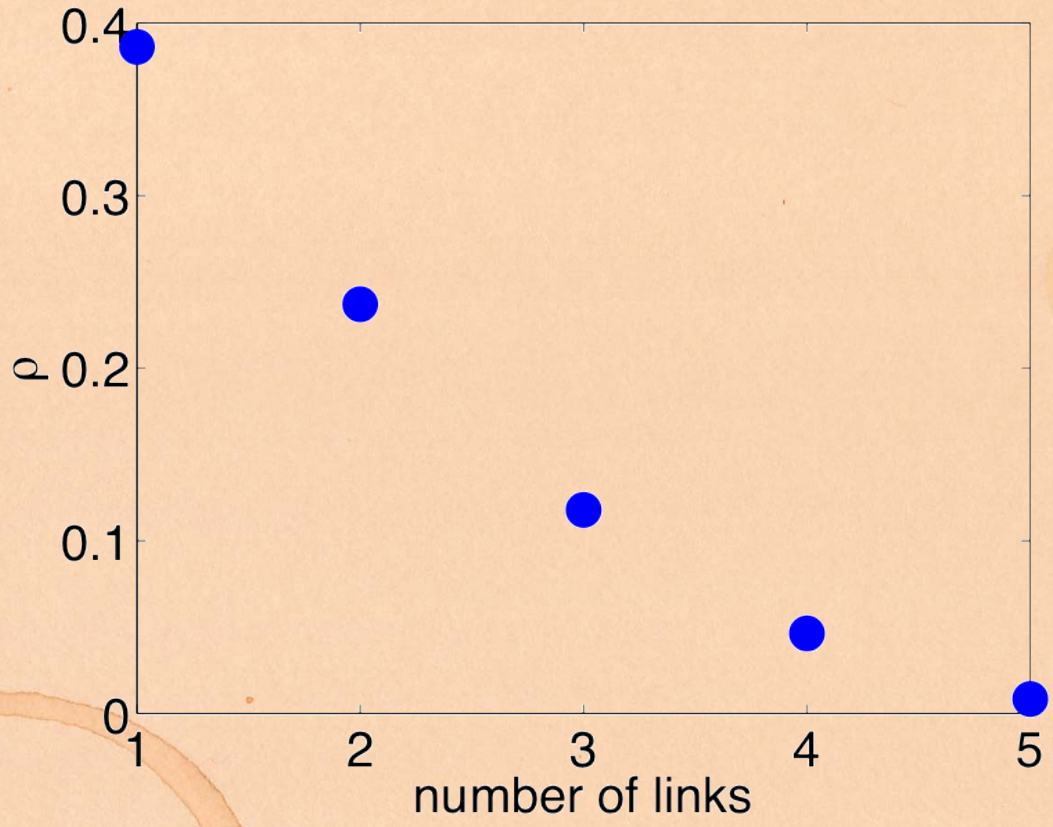
Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References





Twitter—interactions:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References



- Decay in happiness correlation in social network.
- ρ = Spearman's correlation coefficient.

Mechanical Turk action:

valence rank	word	frequency rank	valence	st. dev.
1	happiness	2810	8.44	0.972269
2	love	36	8.42	1.108225
3	happy	96	8.30	0.994885
4	laughed	4948	8.26	1.157231
5	vacation	1482	8.25	1.000000
6	laugh	1559	8.22	1.374550
7	laughing	2420	8.20	1.106567
8	enjoyed	2347	8.18	1.013934
9	celebration	4126	8.18	1.013934
10	excellent	2302	8.18	1.100835
11	congratulations	3379	8.16	1.160943
12	joy	1534	8.16	1.056757
13	successful	2389	8.16	1.075895
14	win	225	8.12	1.081194
15	won	668	8.10	1.216385
16	smile	1443	8.10	1.015191
17	rainbow	4078	8.10	0.994885
18	pleasure	2304	8.08	0.965528
19	winning	2183	8.04	1.049003
20	success	1043	8.02	1.198781
21	award	2188	8.02	1.089530

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Mechanical Turk action:

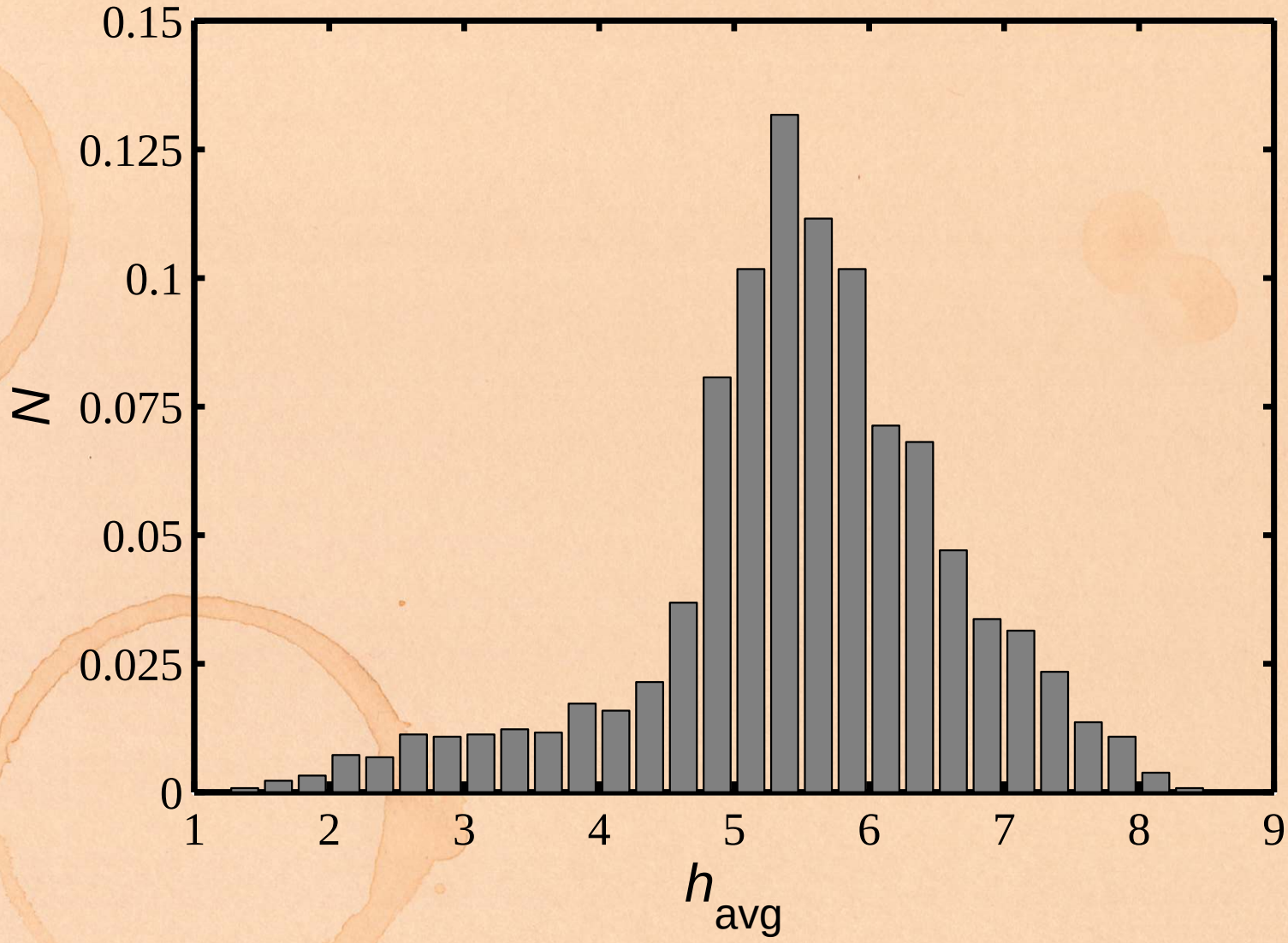
valence rank	word	frequency frequency	valence	st. dev.
4975	failed	1961	1.84	0.997139
4976	cruel	4424	1.84	1.149268
4977	war	440	1.80	1.414214
4978	jail	2513	1.80	0.999575
4979	kills	3686	1.78	1.233710
4980	die	649	1.74	1.191980
4981	killing	2317	1.70	1.359021
4982	arrested	2197	1.67	0.987162
4983	deaths	4451	1.64	1.138563
4984	torture	4712	1.58	1.051529
4985	death	565	1.57	1.274755
4986	died	298	1.56	1.197957
4987	kill	1247	1.56	1.052887
4988	killed	1269	1.56	1.231558
4989	cancer	1240	1.54	1.073046
4990	terrorism	4761	1.51	0.892619
4991	murder	2352	1.48	1.014990
4992	rape	4657	1.44	0.786623
4993	suicide	3203	1.33	0.833688
4994	terrorist	4527	1.30	0.909137

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Variable words:

std dev rank	word	frequency frequency	valence	st. dev.
1	fu[redacted]ing	694	4.64	2.926027
2	fu[redacted]in	1678	3.86	2.740550
3	fu[redacted]ed	2790	3.63	2.690213
4	fu[redacted]	495	4.14	2.579432
5	pu[redacted]y	3051	5.00	2.526456
6	porn	2735	4.18	2.430168
7	beer	1312	5.92	2.389091
8	aids	1845	4.28	2.347730
9	crazy	594	4.64	2.256600
10	drunk	1564	3.88	2.246448
11	drama	2645	5.22	2.243267
12	alcohol	4159	5.31	2.219272
13	prayer	3897	6.48	2.206114
14	chilling	4558	4.76	2.181181
15	tobacco	4467	3.48	2.178185
16	beef	4294	5.28	2.176310
17	rainy	3808	5.64	2.173683
18	raining	3033	5.92	2.165028
19	obama	765	5.94	2.160971
20	walmart	4206	5.45	2.160837
21	palin	2830	4.16	2.151032
22	christ	3759	6.29	2.150581
23	haiti	1211	4.14	2.150581
24	naked	2024	6.02	2.145633
25	payments	3272	5.76	2.143428

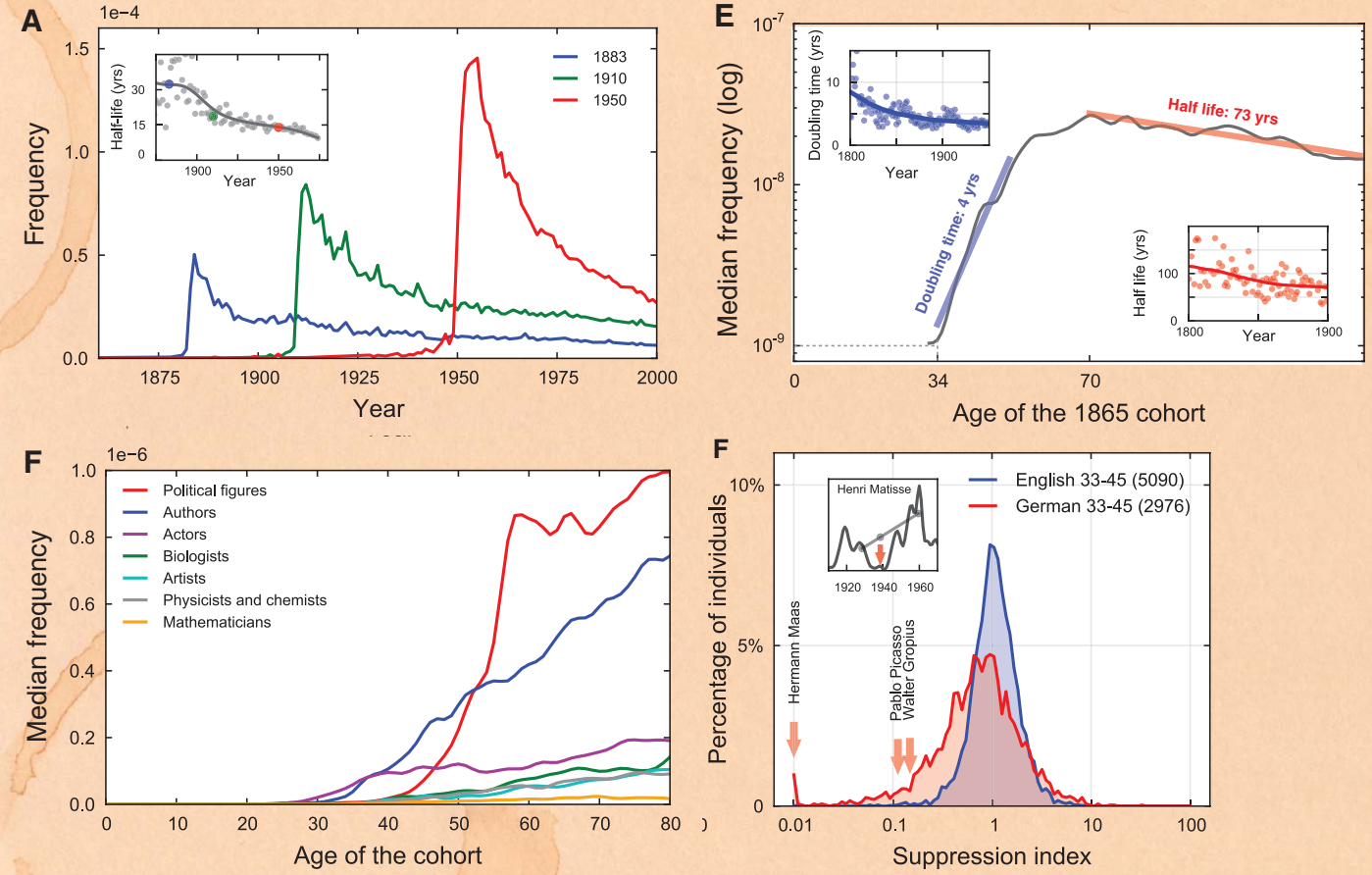
Positive bias in the English language:



Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Culturomics:

"Quantitative analysis of culture using millions of digitized books" by Michel et al., Science, 2011 [13]



http://www.culturomics.org/ (⊞)

Google Books ngram viewer (⊞)

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The University of Vermont

What matters and what's measurable:

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

The UNIVERSITY of VERMONT

Economics, Schmeconomics

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Alan Greenspan (September 18, 2007):

"I've been dealing with these big mathematical models of forecasting the economy ...

If I could figure out a way to determine whether or not people are more fearful or changing to more euphoric,

I don't need any of this other stuff.

I could forecast the economy better than any way I know."

http://wikipedia.org

Economics, Schmeconomics

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

Greenspan continues:

"The trouble is that we can't figure that out. I've been in the forecasting business for 50 years. I'm no better than I ever was, and nobody else is. Forecasting 50 years ago was as good or as bad as it is today. And the reason is that human nature hasn't changed. We can't improve ourselves."

Jon Stewart:

"You just bummed the @*!# out of me."

wildbluffmedia.com

- From the Daily Show (⊞) (September 18, 2007)
- The full inteview is here (⊞).

For more...

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

- PSD, KDH, IMK, CAB, and CMD
  "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter."
  http://arxiv.org/abs/1101.5120 (⊞)
- P. S. Dodds and C. M. Danforth
  "Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents." [7]
  Journal of Happiness Studies, 2009.
- http://www.uvm.edu/∼pdodds/research/ (⊞)
- "Does a Nation's Mood Lurk in Its Songs and Blogs?" by Benedict Carey
  New York Times, August 2009. (⊞)
- http://www.onehappybird.com (⊞)

References I

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

[1] http://www.gutenberg.org/.

[2] http://www.cs.cmu.edu/~enron/.

[3] 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 (⊞)

[4] 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.

The UNIVERSITY of VERMONT

References II

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

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

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

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

References III

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

[9] 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 (⊞)

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

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

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

References IV

Some motivation
Measuring emotional content
Data sets
Analysis
Songs
Blogs
SOTU
Twitter
Mechanical Turk action
Prediction
References

[13] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. A. Lieberman.
Quantitative analysis of culture using millions of digitized books.
Science Magazine, 331:176–182, 2011. pdf (⊞)

[14] E. Sandhaus.
The New York Times Annotated Corpus.
Linguistic Data Consortium, Philadelphia, 2008.

The UNIVERSITY of VERMONT