# Unit 2: Dictionary methods

FSU Summer Methods Workshop

Dr. Rochelle Terman

Department of Political Science
University of Chicago

May 2019

Today: Measuring expressed sentiment in documents
Goal: Classify (measure) sentiment in texts
Method: Dictionary methods

Today: Measuring expressed sentiment in documents
Goal: Classify (measure) sentiment in texts
Method: Dictionary methods

Game Plan:

1) Dictionaries
2) Applying dictionaries to text to measure sentiment
3) Applications, interpretation, and pitfalls

Key Terms:

- Dictionary
- Sentiment analysis
- Word weights

# Dictionaries

- Dictionaries are lists of words belonging to a category.

# Dictionaries

```
## 1      2-faced  negative
## 2      2-faces  negative
## 3           a+  positive
## 4     abnormal  negative
## 5      abolish  negative
## 6   abominable  negative
## 7   abominably  negative
## 8    abominate  negative
## 9  abomination  negative
## 10       abort  negative
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.

# Dictionaries

```
## 1      2-faced  negative
## 2      2-faces  negative
## 3           a+  positive
## 4     abnormal  negative
## 5      abolish  negative
## 6   abominable  negative
## 7   abominably  negative
## 8    abominate  negative
## 9  abomination  negative
## 10       abort  negative
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)

# Dictionaries

```
##              word sentiment :
##             <chr>     <chr>
##  1         abacus     trust
##  2        abandon      fear
##  3        abandon  negative
##  4        abandon   sadness
##  5      abandoned     anger
##  6      abandoned      fear
##  7      abandoned  negative
##  8      abandoned   sadness
##  9    abandonment     anger
## 10    abandonment      fear
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)
  - Specific Emotion (Anger, Joy, Sadness)

# Dictionaries

```
##           word sentiment 
##          <chr>     <chr>
##  1      abacus     trust
##  2     abandon      fear
##  3     abandon  negative
##  4     abandon   sadness
##  5   abandoned     anger
##  6   abandoned      fear
##  7   abandoned  negative
##  8   abandoned   sadness
##  9 abandonment     anger
## 10 abandonment      fear
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)
  - Specific Emotion (Anger, Joy, Sadness)
  - Not all English words included (many are neutral)

# Dictionaries

```
## 1      2-faced  negative
## 2      2-faces  negative
## 3           a+  positive
## 4     abnormal  negative
## 5      abolish  negative
## 6   abominable  negative
## 7   abominably  negative
## 8    abominate  negative
## 9  abomination  negative
## 10       abort  negative
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)
  - Specific Emotion (Anger, Joy, Sadness)
  - Not all English words included (many are neutral)
- Word weights / scores
  - Binary: $\{$Positive $(+1)$, Negative (-1)$\}$

# Dictionaries

```
## 1      abandon   -2
## 2    abandoned   -2
## 3     abandons   -2
## 4     abducted   -2
## 5    abduction   -2
## 6   abductions   -2
## 7        abhor   -3
## 8     abhorred   -3
## 9    abhorrent   -3
## 10      abhors   -3
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)
  - Specific Emotion (Anger, Joy, Sadness)
  - Not all English words included (many are neutral)
- Word weights / scores
  - Binary: {Positive (+1), Negative (-1)}
  - Numerical: $\{-2, -1, 1, 2\}$

# Dictionaries

```
## 1     abandon   -2
## 2   abandoned   -2
## 3    abandons   -2
## 4    abducted   -2
## 5   abduction   -2
## 6  abductions   -2
## 7       abhor   -3
## 8    abhorred   -3
## 9   abhorrent   -3
## 10     abhors   -3
```

- Dictionaries are lists of words belonging to a category.
- Sentiment dictionaries: Lists of words scored according to some emotion or opinion content of the words.
  - Broad Sentiment (Positive, Negative)
  - Specific Emotion (Anger, Joy, Sadness)
  - Not all English words included (many are neutral)
- Word weights / scores
  - Binary: {Positive (+1), Negative (-1)}
  - Numerical: $\{-2, -1, 1, 2\}$
- Non-sentiment dictionaries: Words about sports, food, places...

# Off the Shelf Dictionaries

1) General Inquirer Database (`http://www.wjh.harvard.edu/~inquirer/`)

# Off the Shelf Dictionaries

1) General Inquirer Database
   (`http://www.wjh.harvard.edu/~inquirer/` )
2) Linguistic Inquiry Word Count (LIWC)

# Off the Shelf Dictionaries

1) General Inquirer Database (`http://www.wjh.harvard.edu/~inquirer/` )
2) Linguistic Inquiry Word Count (LIWC)
3) Harvard-IV-4

# Off the Shelf Dictionaries

1) General Inquirer Database (`http://www.wjh.harvard.edu/~inquirer/` )
2) Linguistic Inquiry Word Count (LIWC)
3) Harvard-IV-4
4) Affective Norms for English Words (ANEW)

# Off the Shelf Dictionaries

1) General Inquirer Database (`http://www.wjh.harvard.edu/~inquirer/` )
2) Linguistic Inquiry Word Count (LIWC)
3) Harvard-IV-4
4) Affective Norms for English Words (ANEW)
5) Proprietary software (e.g. Diction)

# Off the Shelf Dictionaries

1) General Inquirer Database (`http://www.wjh.harvard.edu/~inquirer/` )
2) Linguistic Inquiry Word Count (LIWC)
3) Harvard-IV-4
4) Affective Norms for English Words (ANEW)
5) Proprietary software (e.g. Diction)
6) Many many more....

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation

- Careful thought (prayer? epiphanies? divine intervention?) about useful words

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation

- Careful thought (prayer? epiphanies? divine intervention?) about useful words
- Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation
  - Careful thought (prayer? epiphanies? divine intervention?) about useful words
  - Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation
   - Careful thought (prayer? epiphanies? divine intervention?) about useful words
   - Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks
   a) Undergraduates: Pizza $\rightarrow$ Research Output

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation

- Careful thought (prayer? epiphanies? divine intervention?) about useful words
- Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks

a) Undergraduates: Pizza $\rightarrow$ Research Output
b) Mechanical Turkers

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation
  - Careful thought (prayer? epiphanies? divine intervention?) about useful words
  - Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks
  a) Undergraduates: Pizza $\rightarrow$ Research Output
  b) Mechanical Turkers
    - Example: { Happy, Unhappy }

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation
  - Careful thought (prayer? epiphanies? divine intervention?) about useful words
  - Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks
  a) Undergraduates: Pizza $\rightarrow$ Research Output
  b) Mechanical Turkers
    - Example: { Happy, Unhappy }
    - Ask turkers: how happy is `elevator`, `car`, `pretty`, `young`

# Generating Dictionaries

Three ways to create dictionaries (non-exhaustive):

1.) Statistical methods (Distinctive / separating words, next class)

2.) Manual generation
  - Careful thought (prayer? epiphanies? divine intervention?) about useful words
  - Ex: LIWC: " We drew on common emotion rating scales...Roget's Thesaurus...standard English dictionaries. [then] brain-storming sessions among 3-6 judges were held" to generate other words

3.) Crowdsourcing: Populations of people who are surprisingly willing to perform ill-defined tasks
  a) Undergraduates: Pizza → Research Output
  b) Mechanical Turkers
    - Example: { Happy, Unhappy }
    - Ask turkers: how happy is `elevator`, `car`, `pretty`, `young`
    - Output as dictionary

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
    - $\theta_p \in \{0, 1\}$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

For each document $i$ calculate score for document

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
    - $\theta_p \in \{0, 1\}$
    - $\theta_p \in \{-1, 0, 1\}$
    - $\theta_p \in \{-2, -1, 0, 1, 2\}$
    - $\theta_p \in \Re$

For each document $i$ calculate score for document

$$Y_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

For each document $i$ calculate score for document

$$Y_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

$Y_i \approx$ continuous $\rightsquigarrow$ Classification

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

For each document $i$ calculate score for document

$$Y_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

$Y_i \approx$ continuous $\rightsquigarrow$ Classification

$Y_i > 0 \Rightarrow$ Positive Category

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

For each document $i$ calculate score for document

$$Y_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

$Y_i \approx$ continuous $\rightsquigarrow$ Classification

$Y_i > 0 \Rightarrow$ Positive Category

$Y_i < 0 \Rightarrow$ Negative Category

# Applying Dictionaries to Documents

- Vector of word counts: $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})$, $i = (1, \ldots, N)$
- Weights attached to words $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$
  - $\theta_p \in \{0, 1\}$
  - $\theta_p \in \{-1, 0, 1\}$
  - $\theta_p \in \{-2, -1, 0, 1, 2\}$
  - $\theta_p \in \Re$

For each document $i$ calculate score for document

$$Y_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

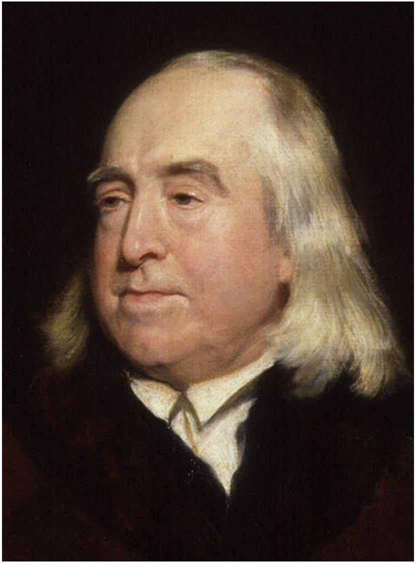
$Y_i \approx$ continuous $\rightsquigarrow$ Classification

$Y_i > 0 \Rightarrow$ Positive Category

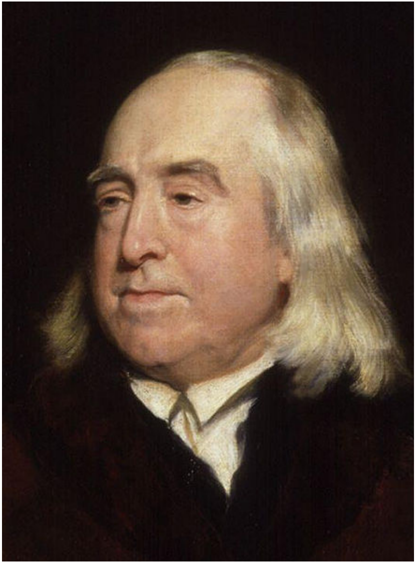
$Y_i < 0 \Rightarrow$ Negative Category

$Y_i \approx 0$ Ambiguous

# Measuring Happiness

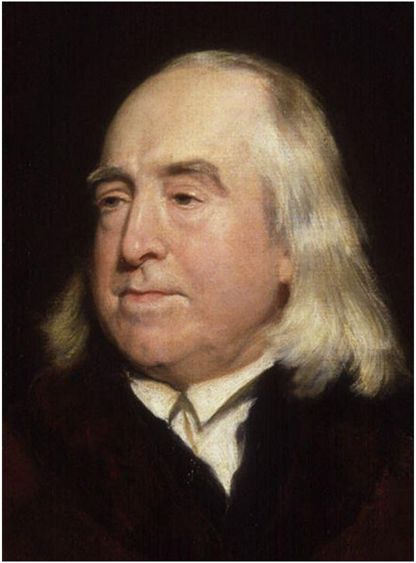


# Measuring Happiness



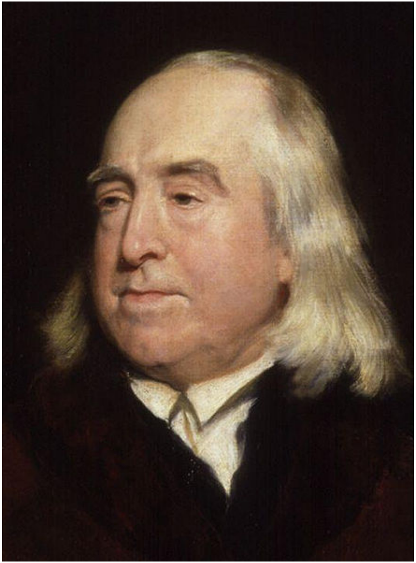
- Quantifying Happiness: How happy is society?

# Measuring Happiness



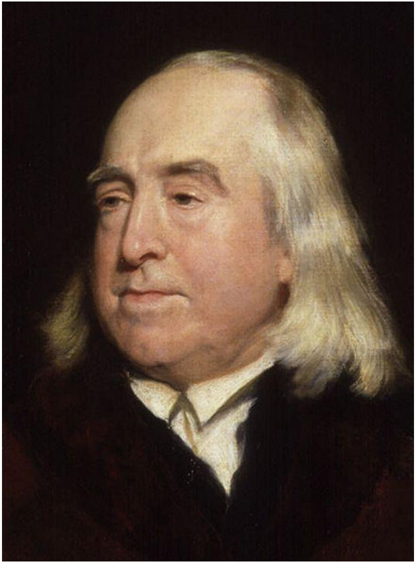
- Quantifying Happiness: How happy is society?
- How Happy is a Song?

# Measuring Happiness



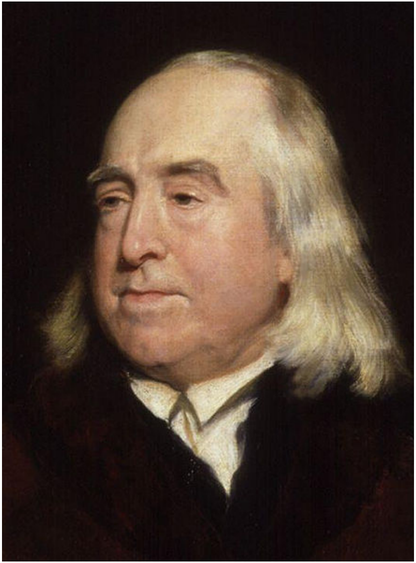
- Quantifying Happiness: How happy is society?
- How Happy is a Song?
- Blog posts?

# Measuring Happiness



- Quantifying Happiness: How happy is society?
- How Happy is a Song?
- Blog posts?
- Facebook posts? (Gross National Happiness)

# Measuring Happiness



- Quantifying Happiness: How happy is society?
- How Happy is a Song?
- Blog posts?
- Facebook posts? (Gross National Happiness)

Use Dictionary Methods

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?
    Happy : triumphant (8.82)/paradise (8.72)/ love (8.72)

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?
    Happy : triumphant (8.82)/paradise (8.72)/ love (8.72)
    Neutral: street (5.22)/ paper (5.20)/ engine (5.20)

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?
    Happy : triumphant (8.82)/paradise (8.72)/ love (8.72)
    Neutral: street (5.22)/ paper (5.20)/ engine (5.20)
    Unhappy : cancer (1.5)/funeral (1.39)/ rape (1.25) /suicide (1.25)

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?
    Happy : triumphant (8.82)/paradise (8.72)/ love (8.72)
    Neutral: street (5.22)/ paper (5.20)/ engine (5.20)
    Unhappy : cancer (1.5)/funeral (1.39)/ rape (1.25) /suicide (1.25)
- Happiness for text $i$ (with word $p$ having happiness $\theta_p$ and document frequence $X_{ip}$

# Measuring Happiness

Dodds and Danforth (2009): Use a dictionary method to measure happiness

- Affective Norms for English Words (ANEW)
- Bradley and Lang 1999: 1034 words, Affective reaction to words
  - On a scale of 1-9 how happy does this word make you?
    Happy : triumphant (8.82)/paradise (8.72)/ love (8.72)
    Neutral: street (5.22)/ paper (5.20)/ engine (5.20)
    Unhappy : cancer (1.5)/funeral (1.39)/ rape (1.25) /suicide (1.25)
- Happiness for text $i$ (with word $p$ having happiness $\theta_p$ and document frequence $X_{ip}$

$$\text{Happiness}_i = \frac{\sum_{p=1}^{P} \theta_p X_{ip}}{\sum_{p=1}^{P} X_{ip}}$$

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.
⋮

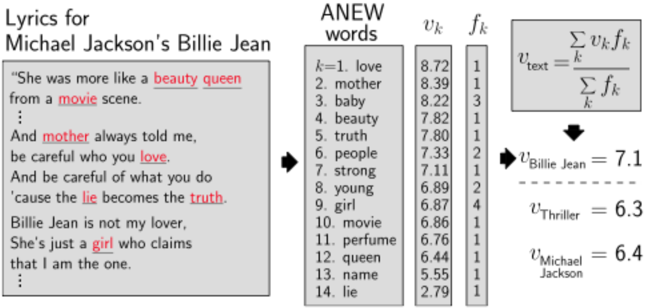
| ANEW words | $v_k$ | $f_k$ |
|---|---|---|
| $k$=1. love | 8.72 | 1 |
| 2. mother | 8.39 | 1 |
| 3. baby | 8.22 | 3 |
| 4. beauty | 7.82 | 1 |
| 5. truth | 7.80 | 1 |
| 6. people | 7.33 | 2 |
| 7. strong | 7.11 | 1 |
| 8. young | 6.89 | 2 |
| 9. girl | 6.87 | 4 |
| 10. movie | 6.86 | 1 |
| 11. perfume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

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

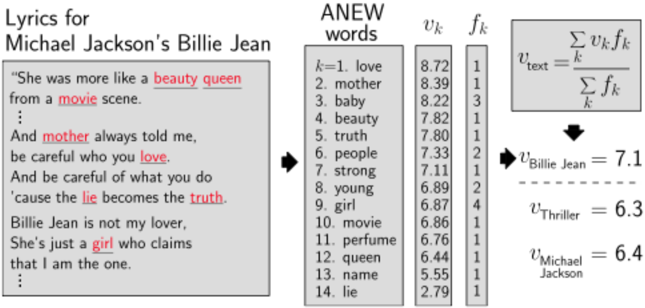
$$v_{\text{Billie Jean}} = 7.1$$

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

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



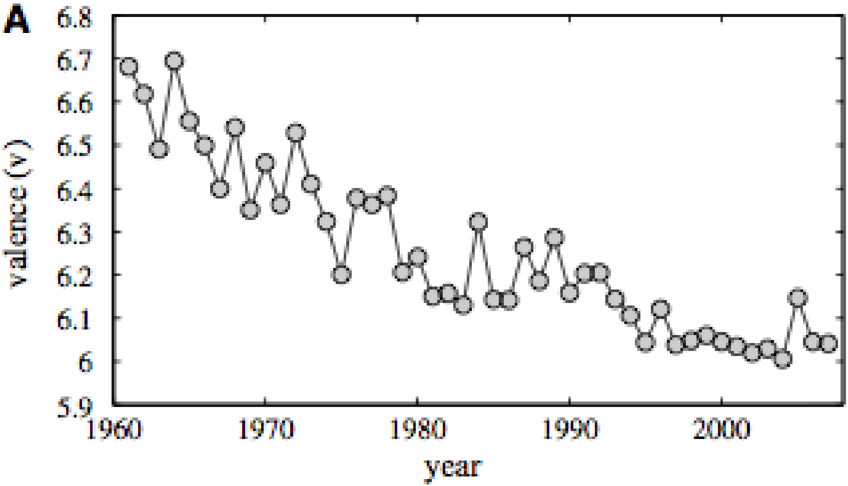

Happiest Song on Thriller?



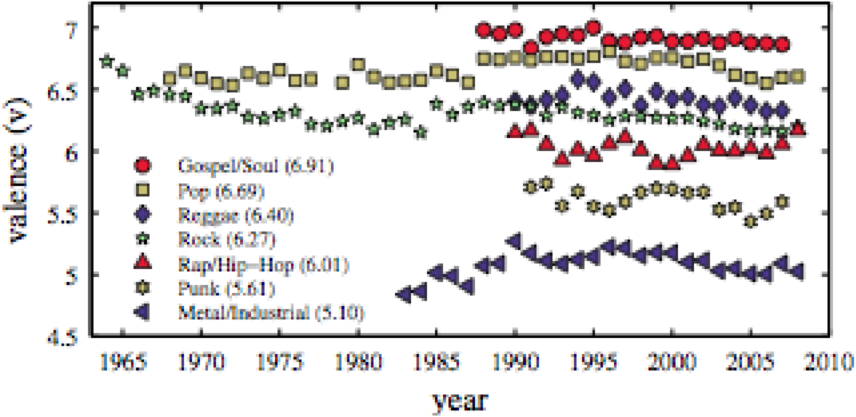

Happiest Song on Thriller?

P.Y.T. (Pretty Young Thing) (This is the right answer!)

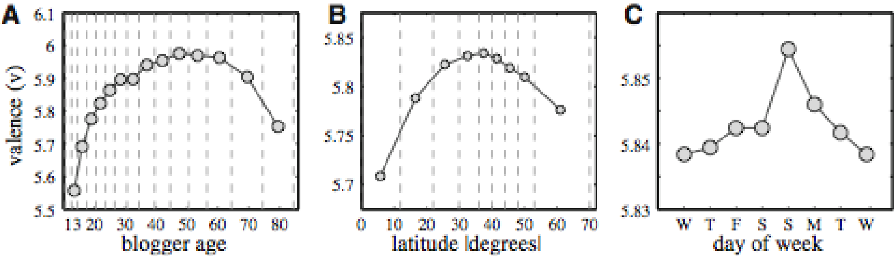
# Happiness in Society



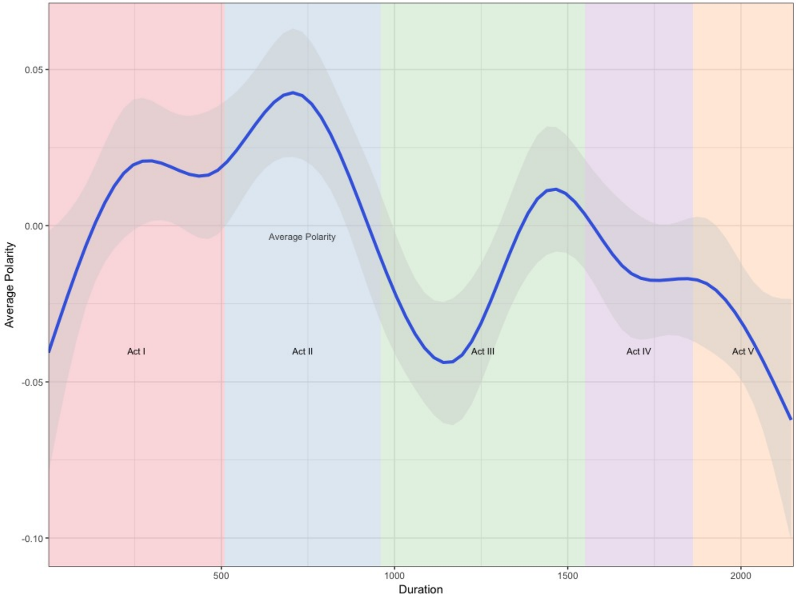
# Happiness in Society



# Happiness in Society



# Visualizing Plots: Romeo & Juliet



# Emotional Contagion on Facebook

www.pnas.org

**Experimental evidence of massive-scale emotional contagion through social networks**

# Emotional Contagion on Facebook

# Use with Caution!

- Most dictionary methods do not take into account qualifiers (e.g. “no good”).

# Use with Caution!

- Most dictionary methods do not take into account qualifiers (e.g. "no good").
- Ignores sarcasm, irony, nuance.

# Use with Caution!

- Most dictionary methods do not take into account qualifiers (e.g. "no good").
- Ignores sarcasm, irony, nuance.
- Dictionaries are context dependent.

# Use with Caution!

- Most dictionary methods do not take into account qualifiers (e.g. "no good").
- Ignores sarcasm, irony, nuance.
- Dictionaries are context dependent.

# Validation, Dictionaries from other Fields

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting:

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting: `tax, cost, capital, board, liability, foreign, cancer, crude (oil), tire`

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting: `tax, cost, capital, board, liability, foreign, cancer, crude (oil), tire`
- 73% of Harvard negative words in this set(!!!!!)

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting: `tax, cost, capital, board, liability, foreign, cancer, crude (oil), tire`
- 73% of Harvard negative words in this set(!!!!!)
- Not Negative Harvard, Negative in Accounting:

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting:
  `tax, cost, capital, board, liability, foreign, cancer, crude (oil), tire`
- 73% of Harvard negative words in this set(!!!!!)
- Not Negative Harvard, Negative in Accounting:
  `felony, litigation, restated, misstatement, unanticipated`

# Validation, Dictionaries from other Fields

Accounting Research: measure tone of 10-K reports

- tone matters ($)

Previous state of art: Harvard-IV-4 Dictionary applied to texts

Loughran and McDonald (2011): Financial Documents are Different, polysemes

- Negative words in Harvard, Not Negative in Accounting: `tax`, `cost`, `capital`, `board`, `liability`, `foreign`, `cancer`, `crude (oil)`, `tire`
- 73% of Harvard negative words in this set(!!!!!)
- Not Negative Harvard, Negative in Accounting: `felony`, `litigation`, `restated`, `misstatement`, `unanticipated`

Context Matters

`R Code!`