

# Visualizing Text

Yannet Interian -- USF

# Agenda

- Example: Health Care Reform
- Text as Data
- Visualizing Document
- Content Evolving Documents
- Visualizing Conversation
- Document Collections
- Other cool examples of Text Viz

# Why Visualize Text?

- Understanding – get the “gist” of a document
- Grouping – cluster for overview or classification
- Comparison – compare document collections, or inspect evolution of collection over time
- Correlation – compare patterns in text to those in other data, e.g., correlate with social network

# Text as Data

- Documents
  - Articles, books and novels
  - E-mails, web pages, blogs
  - Tags, comments
  - Computer programs, logs
- Collections of Documents
  - Messages (e-mail, blogs, tags, comments)
  - Social networks (personal profiles)
  - Academic collaborations (publications)

# Example: Health Care Reform

- Recent History
  - Initiatives by President Clinton
  - Overhaul by President Obama
- Text Data
  - News articles
  - Speech transcriptions
  - Legal documents
- What questions might you want to answer?
- What visualizations might help?

# A Concrete Example

September 10, 2009

TEXT

## **Obama's Health Care Speech to Congress**

Following is the prepared text of President Obama's speech to Congress on the need to overhaul health care in the United States, as released by the White House.

Madame Speaker, Vice President Biden, Members of Congress, and the American people:

When I spoke here last winter, this nation was facing the worst economic crisis since the Great Depression. We were losing an average of 700,000 jobs per month. Credit was frozen. And our financial system was on the verge of collapse.

As any American who is still looking for work or a way to pay their bills will tell you, we are by no means out of the woods. A full and vibrant recovery is many months away. And I will not let up until those Americans who seek jobs can find them; until those businesses that seek capital and credit can thrive; until all responsible homeowners can stay in their homes. That is our ultimate goal. But thanks to the bold and decisive action we have taken since January, I can stand here with confidence and say that we have pulled this economy back from the brink.

I want to thank the members of this body for your efforts and your support in these last several months, and especially those who have taken the difficult votes that have put us on a path to recovery. I also want to thank the American people for their patience and resolve during this trying time for our nation.

But we did not come here just to clean up crises. We came to build a future. So tonight, I return to speak to all of yo

# Tag Clouds: Word Count

## President Obama's Health Care Speech to Congress [NY Times]



**people will care can Americans believe one way achieve plans give doctor paperwork taken Now even medicine**

**every system using insurance hospital coverage proposal better going sick child thousands third may**

**right business continue government still children plan simply also following**

**public budget companies understand hospitals never employer nation someone**

**choice best money American doctors first years think principles get long**

**whether work journey now percent small go cost agree last let like keep cover responsibility country thing pay make savings**

**something reform inflation else Congress less ask arguments look many change much covered**

**higher able just time security premiums employees today quality ever businesses services**

**new elderly everybody problem tonight disagreed**

**family**

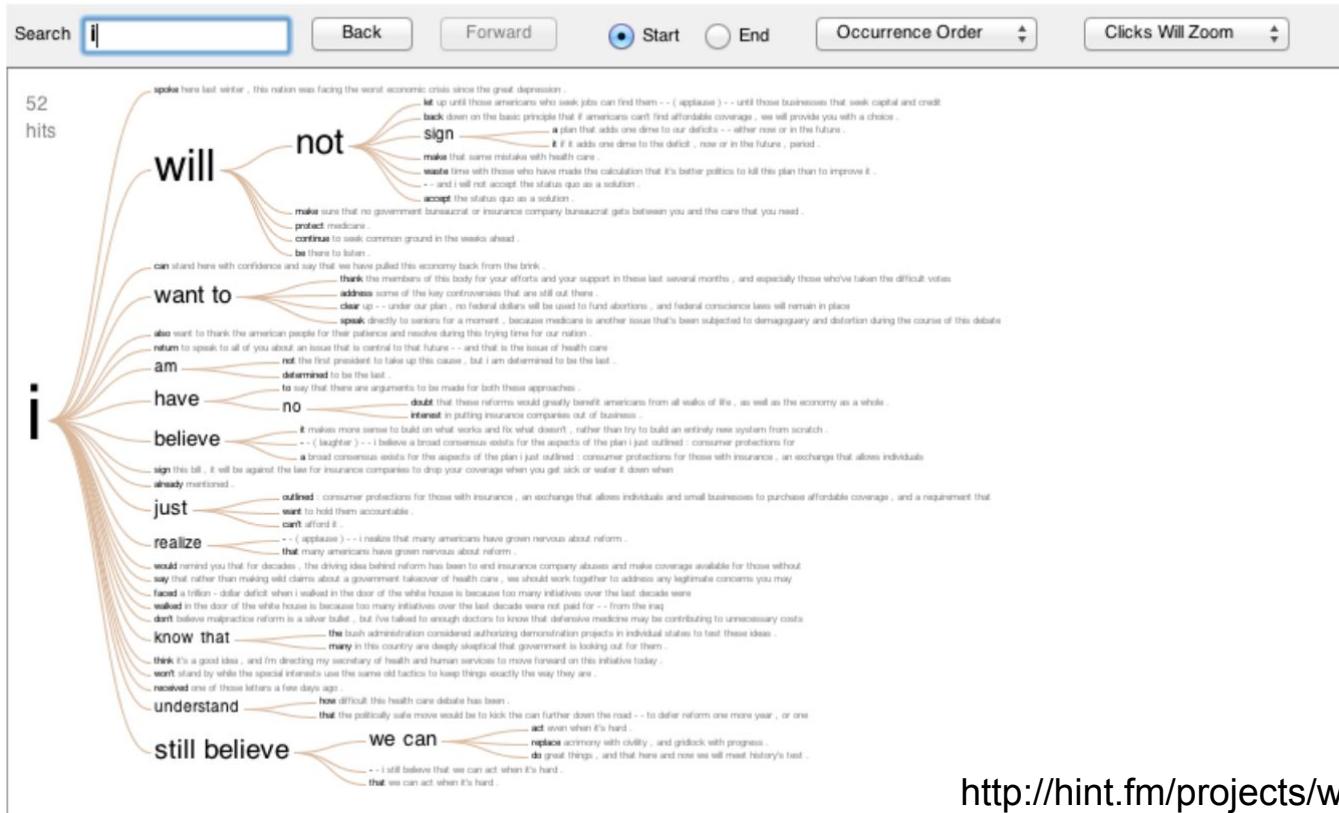
Bill Clinton 1993

# Barack Obama 2009

millions  
nothing three still exchange Now meet things many believe business  
coverag greater come care know time proposing  
small country option must back savings change always customers  
pay without us friend people costs want  
one afford make nation company provide deficit doctors  
right since especially money also competition lose months  
seniors point reform dollar American seen  
public cost seek even Republican stand say  
ow ideas private Congress together last room basic  
ll cost place sick together last may stand  
private Congress sick together last room basic  
Medicare like help offer  
businesses can year America  
Medicare can year America  
insurance just Americans chamber tonight problem debate  
able future already way workers children  
future children waste years chamber tonight problem debate  
individuals medical Medicaid medical

# Word Tree: Word Sequences

## Visualizations : Word Tree President Obama's Address to Congress on Health Care



Search

**i will**

Back

Forward

 Start End

Occurrence Order

Clicks Will Zoom

12  
hits**i will****not**

**let** up until those americans who seek jobs can find them - - ( applause ) - - until tho

**back** down on the basic principle that if americans can't find affordable coverage , w

**sign** a plan that adds one dime to our deficits - - either now or in the future

it if it adds one dime to the deficit , now or in the future , period .

**make** that same mistake with health care .

**waste** time with those who have made the calculation that it's better politics to kill thi

- - and i will not accept the status quo as a solution .

**accept** the status quo as a solution .

**make** sure that no government bureaucrat or insurance company bureaucrat gets between you and t

**protect** medicare .

**continue** to seek common ground in the weeks ahead .

**be** there to listen .

# Evaluation of Text Visualization

Many text visualizations do not represent the text directly. They represent the output of a language model (word counts, word sequences, etc.).

- Can you interpret the visualization? How well does it convey the properties of the model?
- Do you trust the model? How does the model enable us to reason about the text?

# **Text as Data**

# Words as nominal data?

- High dimensional (10,000+)
- More than equality tests
- Words have meanings and relations
  - Correlations: Hong Kong, San Francisco, Bay Area
  - Order: April, February, January, June, March, May
  - Membership: Tennis, Running, Swimming, Hiking, Piano
  - Hierarchy, antonyms & synonyms, entities, ...

# Text Processing Pipeline

1. Tokenization Segment text into terms.
  - a. Remove stop words? a, an, the, of, to, be
  - b. Numbers and symbols? #gocard, @stanfordfbball, Beat Cal!!!!!!!
  - c. Entities? San Francisco, O'Connor, U.S.A.

# Text Processing Pipeline

1. Tokenization Segment text into terms.
  - a. Remove stop words? a, an, the, of, to, be
  - b. Numbers and symbols? #gocard, @stanfordfbball, Beat Cal!!!!!!!
  - c. Entities? San Francisco, O'Connor, U.S.A.
2. Stemming
  - a. Group together different forms of a word.
  - b. Porter stemmer? visualization(s), visualize(s), visually -> visual
  - c. Lemmatization? goes, went, gone -> go
3. Ordered list of terms

# Bag of Words Model

- Ignore ordering relationships within the text
- A document  $\approx$  vector of term weights
  - Each dimension corresponds to a term (10,000+)
  - Each value represents the relevance
    - For example, simple term counts
- Aggregate into a document-term matrix
  - Document vector space model

# Document-Term Matrix

Each document is a vector of term weights

Simplest weighting is to just count occurrences

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	157	73	0	0	0	0
Brutus	4	157	0	1	0	0
Caesar	232	227	0	2	1	1
Calpurnia	0	10	0	0	0	0
Cleopatra	57	0	0	0	0	0
mercy	2	0	3	5	5	1
worser	2	0	1	1	1	0

# Wordle tag cloud for Sarah Palin's Vice Presidential nomination convention speech:



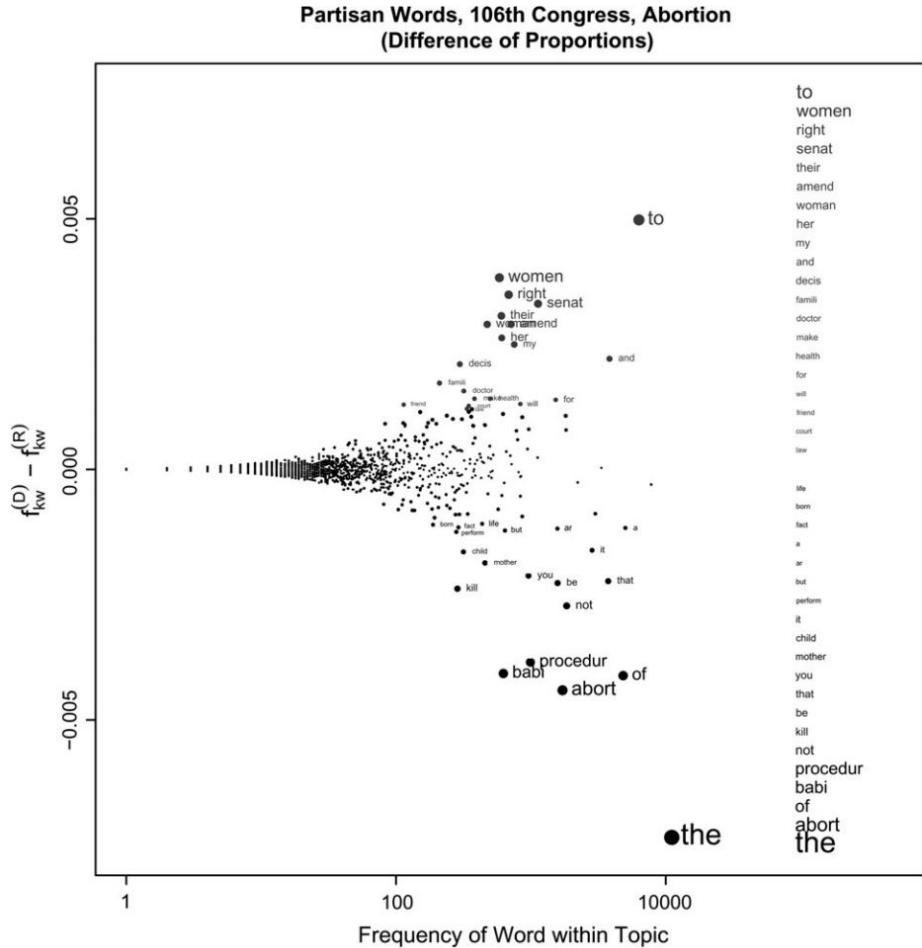
# Tag Clouds

- Strengths
  - Can help with gisting and initial query formation.
- Weaknesses
  - Sub-optimal visual encoding (size vs. position)
  - Inaccurate size encoding (long words are bigger)
  - May not facilitate comparison (unstable layout)
  - Term frequency may not be meaningful
  - Does not show the structure of the text

**Given a text, what are the  
best descriptive words?**

# Difference of Proportions

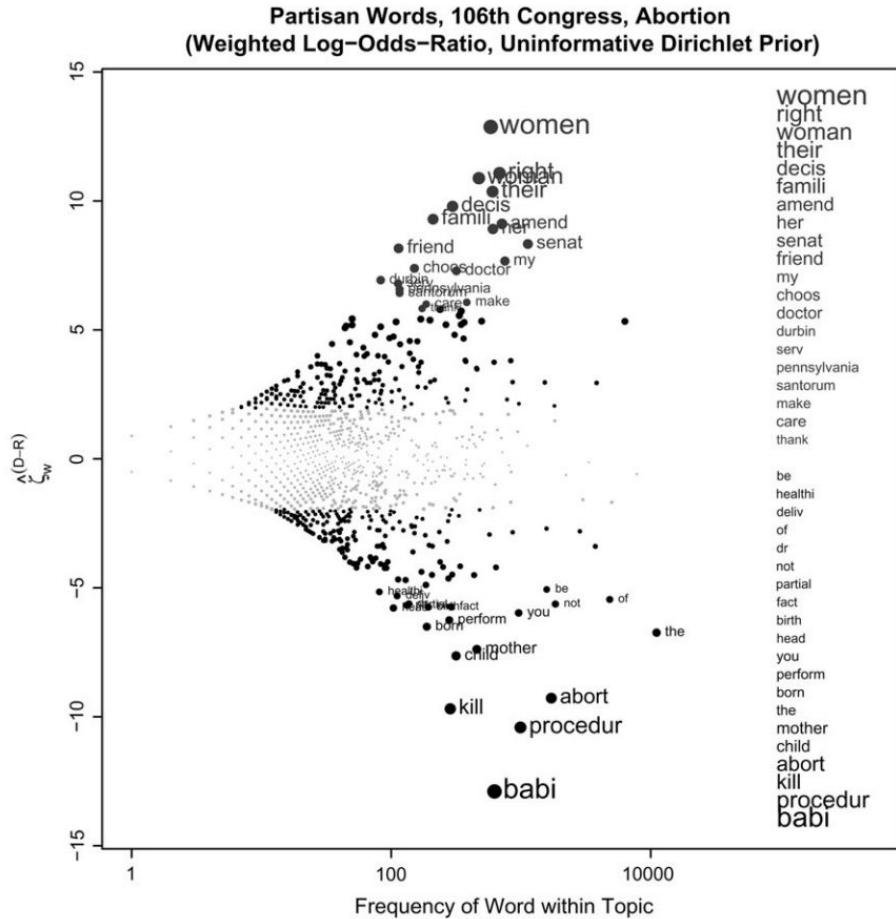
# Partisanship of words on the topic of abortion during the 106th (1999–2000) Senate



# Weighted Log-Odd Ratio

## Model based approach

- Language model of the republican and democratic speech (multinomial)
- Compare the fitted proportions for each word



# Keyword Weighting

## Term Frequency

$$tf_{td} = \text{count}(t) \text{ in } d / \text{count}(d)$$

## TF.IDF: Term Freq by Inverse Document Freq

$$tf.idf_{td} = \times \log(N/df_t)$$

$df_t$  = # docs containing t

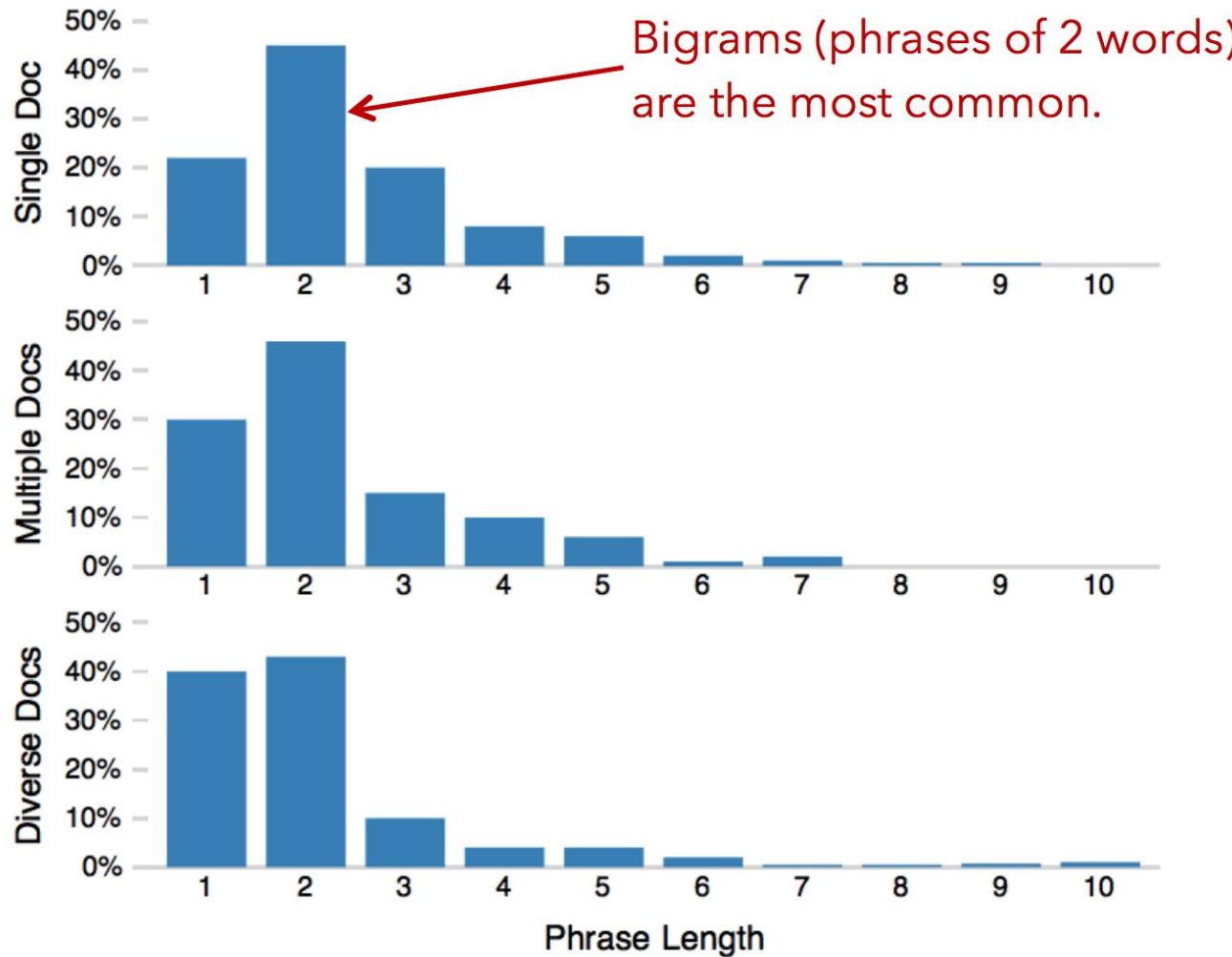
N = # of docs

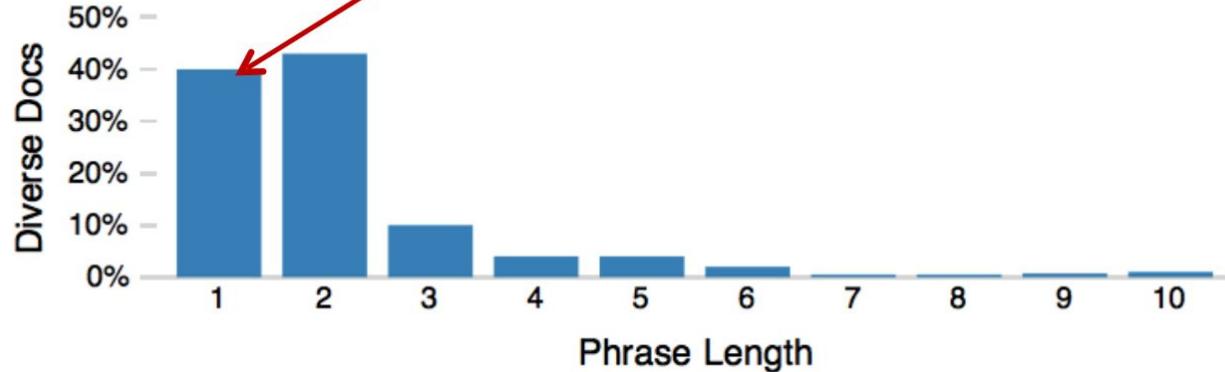
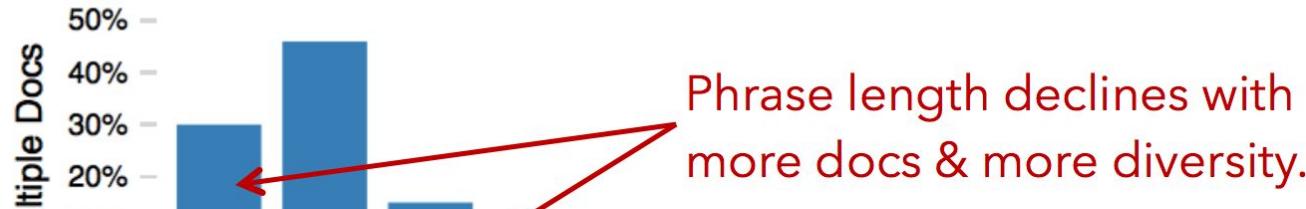
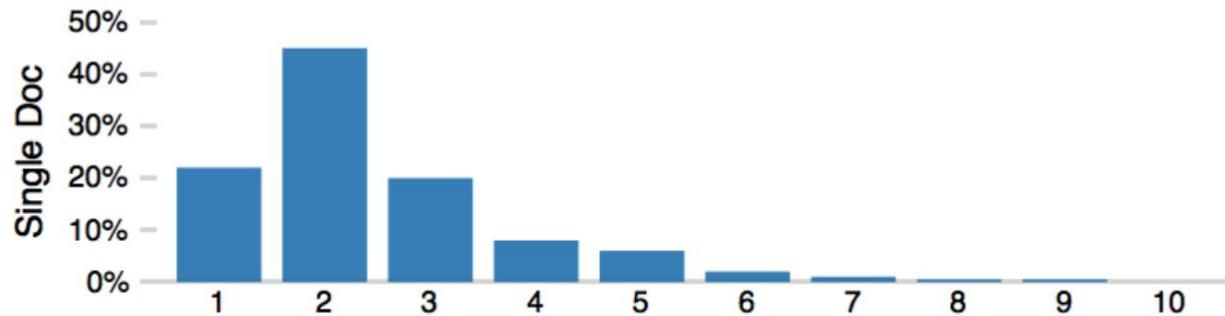
# Limitations of Freq. Statistics

- Typically focus on unigrams (single terms)
- Often favors frequent (TF) or rare (IDF) terms
- A “bag of words” ignores additional information
  - Grammar / part-of-speech
  - Position within document
  - Recognizable entities

# How do people describe text?

- Asked 69 subjects (graduate students) to read and describe dissertation abstracts
- Students were given 3 documents in sequence; they then described the collection as a whole
- Students were matched to both familiar and unfamiliar topics; topical diversity within a collection was varied systematically





# Term Commonness

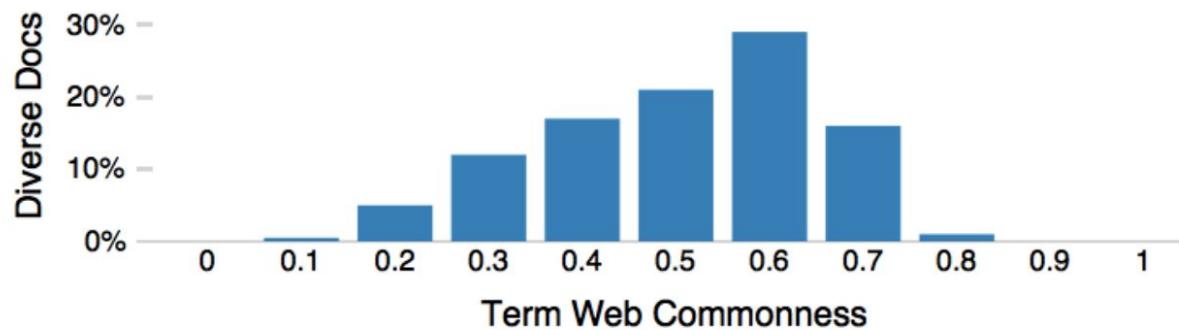
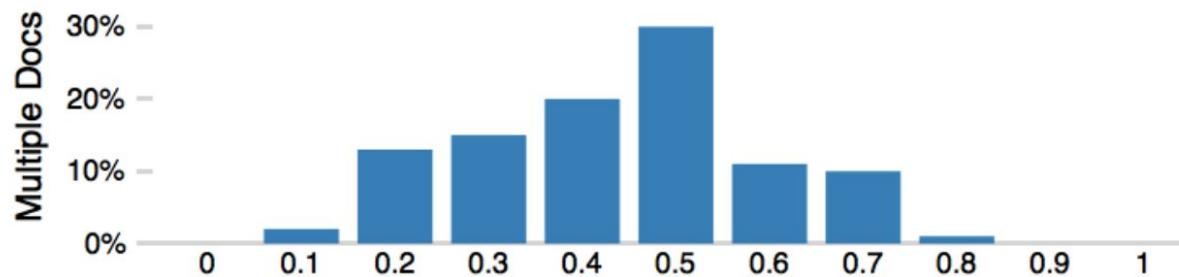
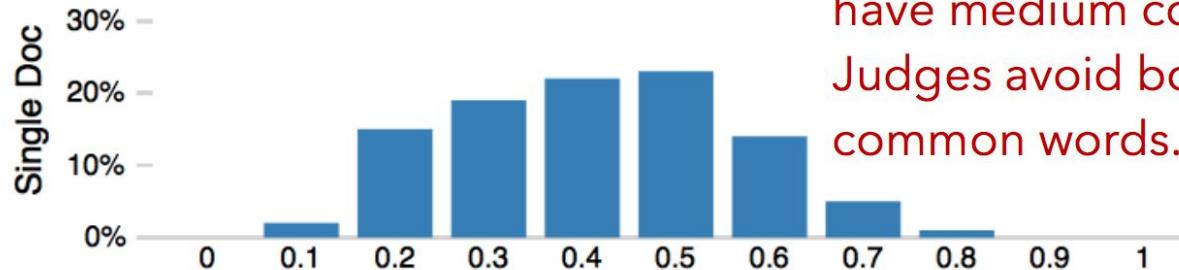
$$tf_w = \text{count}(w)$$

$$\log(tf_w) / \log(tf_{\text{the}})$$

The normalized term frequency relative to the most frequent n-gram, e.g., the word “the”.

Measured across a corpus or across the entire English language (using Google n-grams)

Selected descriptive terms have medium commonness. Judges avoid both rare and common words.



Commonness increases with more docs & more diversity.

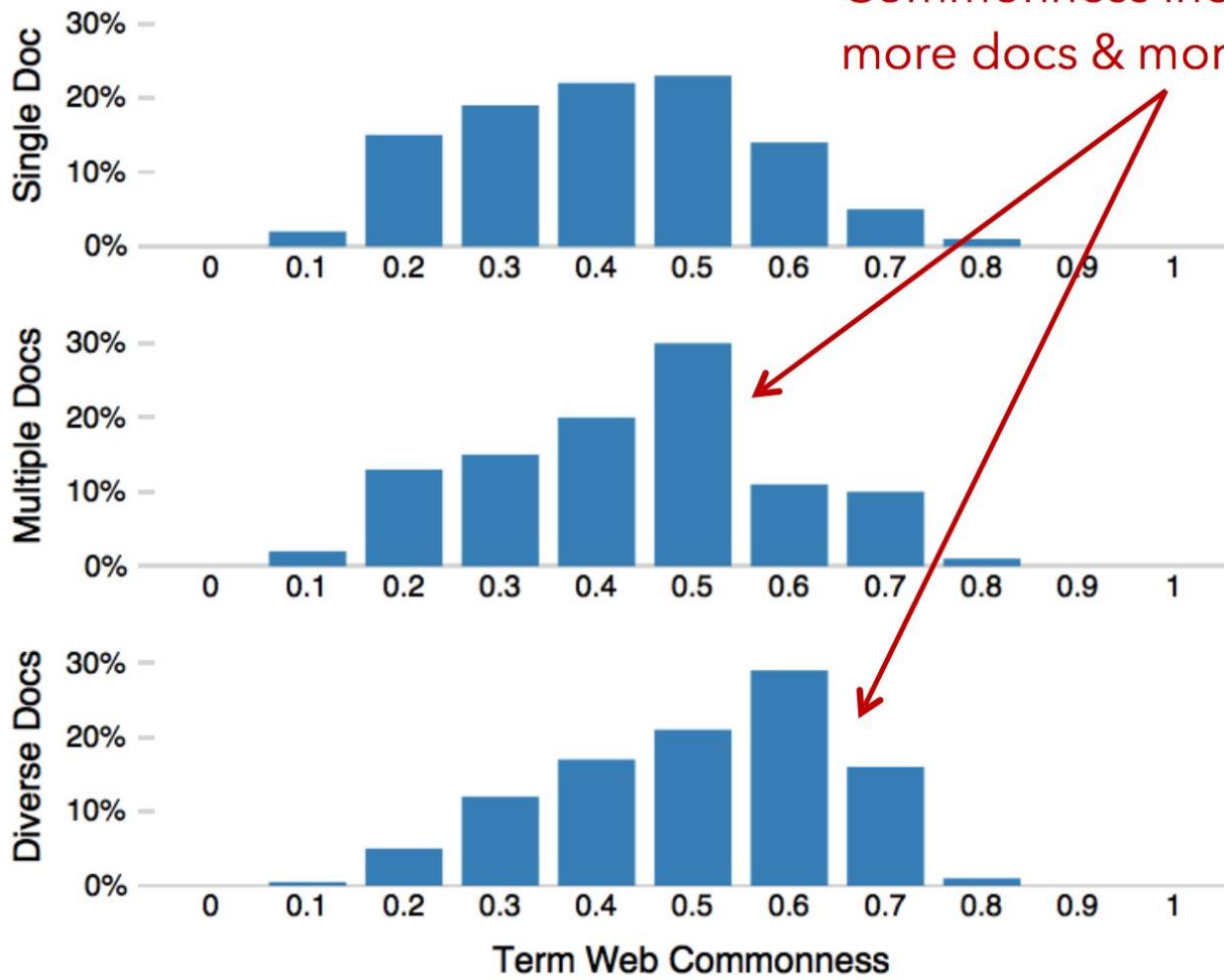


Table IV. Regression Coefficients for the Full (Corpus-Dependent) Model  
Based on the Ph.D. Dissertations

# Model of Descriptive Keyphrases

Model Feature	Regression Coefficients
(intercept)	-2.88114***
log(tf)	0.74095***
WC ∈ (0%, 20%]	0.08894
WC ∈ (20%, 40%]	0.04390
WC ∈ (40%, 60%]	-0.19786
WC ∈ (60%, 80%]	-0.46664*
WC ∈ (80%, 100%]	-1.26714***
CC ∈ (0%, 20%]	0.20554
CC ∈ (20%, 40%]	0.39789**
CC ∈ (40%, 60%]	0.24929
CC ∈ (60%, 80%]	-0.34932
CC ∈ (80%, 100%]	-0.97702**
relative first occurrence	0.52950***
first sentence	0.83637**
partial noun phrase	0.14117
noun phrase	0.29818*
head noun	-0.16509
optional leading word	0.46481*
partial verb phrase	0.15639
verb phrase	1.12310*
full technical term	-0.58959
partial technical term	1.37875*
full compound technical term	1.09713
partial compound technical term	1.10565*

Note: WC = Web commonness, CC = corpus commonness; statistical significance = \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$ .



## A fighter jet rain check

Story and video by [Chamila Jayaweera](#)

Have you ever thought about what it takes to make sure that sea-based fighter jets stay dry?

When it comes to the F/A-18 Super Hornet, Boeing engineers in St. Louis use a special process called the Water Check Test to rule out areas where moisture could seep into the aircraft and its electronics suite.

Program experts douse the jet with simulated rain at a 15-inch-per-hour rate for about 20 minutes inside an enormous hangar in St. Louis.

"Our ultimate customers are U.S. Navy fighter pilots, and we want to ensure their safety in flight and on the ground, and water-tight integrity of the aircraft also helps increase their effectiveness," said Boeing's Rich Baxter, F/A-18 Super Hornet final assembly manager.

To find out moreabout how the process works and watch the action unfold, click above to see the video story.



CHAMILA JAYAWEERA/BOEING

The Water Check team rolls in a large metal frame, which they affectionately call their "spray tree," over a Super Hornet inside a St. Louis hangar.



**fighter**

F/A

Hornet

Super

Boeing  
-18

rain

St.

jet

Louis

15-inch-per-hour  
douse

hangar  
water-tight

Check  
Baxter

sea-based  
aircraft

Rich  
seep

click  
Navy

sure  
Water  
moisture  
watch  
enormous  
stay

**Super Hornet**

F/A -18

**fighter jet**

Boeing engineers  
special process

rain check

electronics suite

Program experts

simulated rain

ultimate customers

enormous hangar  
water-tight integrity

Rich Baxter

15-inch-per-hour rate

video story  
aircraft

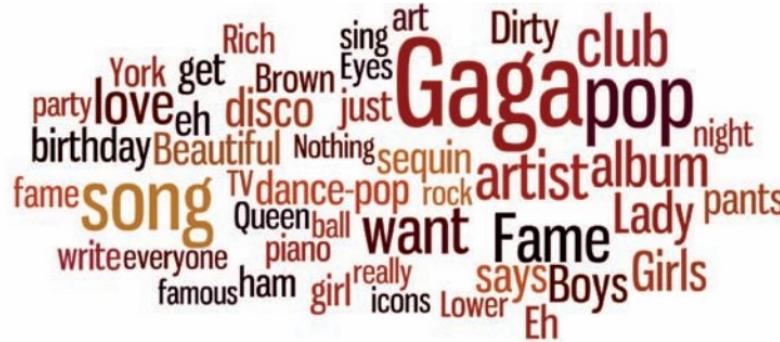
U.S. Navy fighter pilots  
Super Hornet final assembly manager



# Online biography of Lady Gaga

# Single-word phrases (unigrams)

# Multiword phrases using model



# Tips: Descriptive Phrases

- Understand the limitations of your language model.
  - Bag of words:
    - Easy to compute
    - Single words
    - Loss of word ordering
- Select appropriate model and visualization
  - Generate longer, more meaningful phrases
  - Adjective-noun word pairs for reviews
  - Show keyphrases within source text

# Yelp Reviews

'09 amazing around baked bar bass **best** chef delicious eat  
elite everything favorite **fish food fresh** going hamachi  
hawaiian **hour** line love **mango** minutes mussels name  
night nigiri order **people** <sup>prices</sup> really restaurant **roll**  
expensive or cheap?  
sake salmon sea seated service spicy stars sure  
**sushi**  
table think tuna **wait** waitress worth  
“long wait” or “no wait”? what type of sushi roll?

b) best sf  
baked sea bass  
**best sushi**  
sure in striped bass  
other person  
fresh fish slow service  
sushi chef baked mussel more hour  
**sushi bar**  
only thing  
long wait long time  
long line sushi restaurant good food  
hawaiian roll reasonable price  
baked mango  
**small place** delicious everything

**Mentioned 63 times**

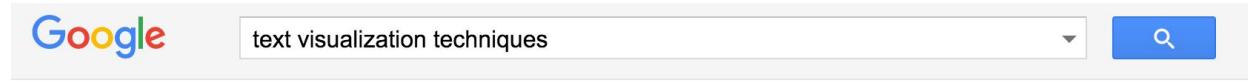
possess sage of the halos wisdom , and know in advance sushi zone only accepts cash  
and the waits will be **long** and arduous .

yes , its a **long** wait , learn the master of zen if you want to eat here .

# Document Content

# Information Retrieval

Search for documents Match query string with documents  
Visualization to **contextualize** results



Scholar      About 1,930,000 results (0.07 sec)

---

Articles      [A taxonomy of visualization techniques using the data state reference model](#)  
[EH Chi - Information Visualization, 2000. InfoVis 2000. IEEE ..., 2000 - ieeexplore.ieee.org](#)  
... Multi-dimensional Tables, Information Landscapes and Spaces, Node and Link, Trees, and **Text** Transforms. OLIVE is a taxonomy assembled by students in Shneiderman's information **visualization** class [Olive99], and divides information **visualization techniques** using eight ...  
Cited by 580   [Related articles](#)   [All 22 versions](#)   [Cite](#)   [Save](#)

---

Any time      [Termite: Visualization techniques for assessing textual topic models](#)  
[J Chuang, CD Manning, J Heer - Proceedings of the International ..., 2012 - dl.acm.org](#)  
... The size of **text** corpora, however, often exceeds the limit of what a person can read and process. ... We present Termite, a **visualization** system for the term- topic distributions produced by topic models. Our system contributes two novel **techniques** to aid topic model assessment. ...  
Cited by 115   [Related articles](#)   [All 10 versions](#)   [Cite](#)   [Save](#)

---

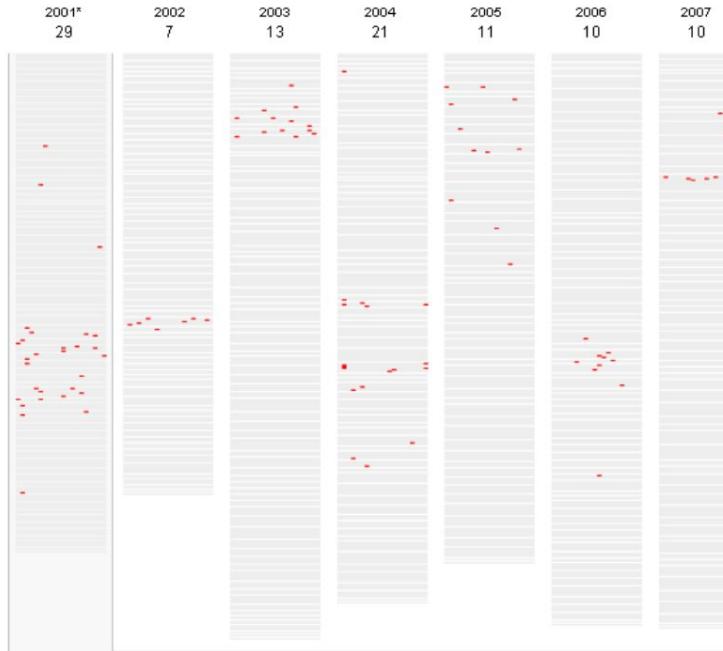
Sort by relevance      [Efficient visualization of large text corpora](#)  
[M Grobelnik, D Mladenić - Proceedings of the seventh TELRI seminar. ..., 2002 - ailab.ijs.si](#)  
... Most frequent application of **text visualization techniques** is particular in cases when one needs to understand or to explain the structure and nature of large quantity of typically ...

# The 2007 State of the Union Address

Over the years, President Bush's State of the Union address has averaged almost 5,000 words each, meaning the the President has delivered over 34,000 words. Some words appear frequently while others appear only sporadically. Use the tools below to analyze what Mr. Bush has said.

  or choose a word here. 

## Use of the phrase "Tax" in past State of the Union Addresses



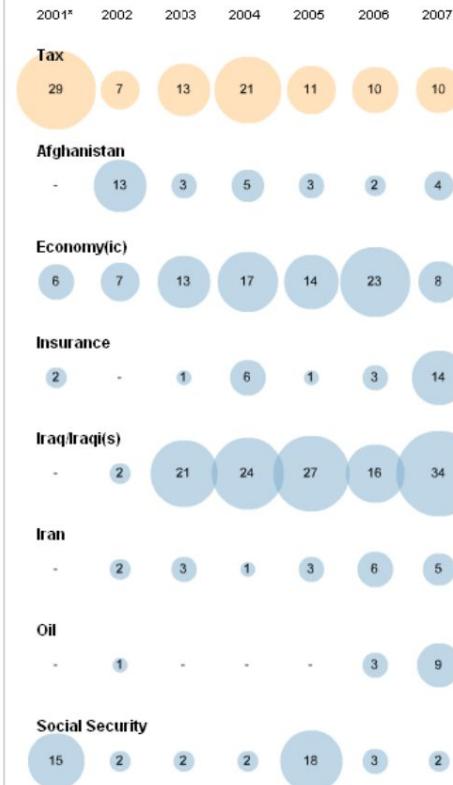
### The word in context

I believe in local control of schools. We should not, and we will not, run public schools from Washington, D.C. Yet when the federal government spends TAX dollars, we must insist on results. Children should be tested on basic reading and math skills every year between grades three and eight. Measuring is the only way to know whether all our children are learning. And I want to know, because I refuse to leave any child behind in America.

-- 2001 (Paragraph 14 of 73)

### Next Instance of 'Tax'

## Compared with other words



\* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

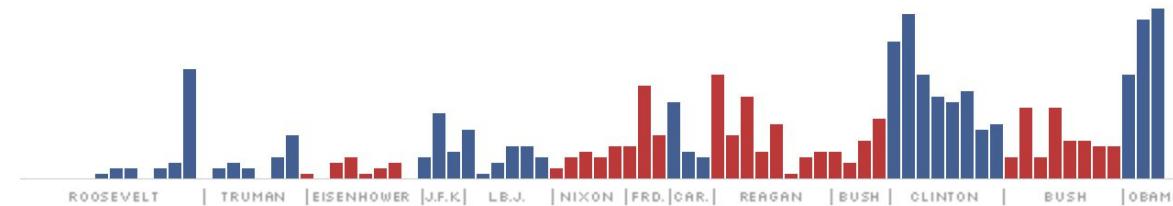
## Patterns of Speech: 75 Years of the State of the Union Addresses

In 2010, President Obama was the first modern president to use the words “bubble,” “supermajority” and “obesity” in a State of the Union speech. But other words have a longer history. Below, a historical look at the number of times presidents have used selected words in their State of the Union addresses (or analogous speeches) from 1934 to 2011.

### ‘jobs’

With unemployment above 9 percent, jobs were a focus of President Obama's speech. Historically, jobs get mentioned in the speech in rough correlation to the economic cycle, with spikes around 1975, 1981, 1991 and 2002.

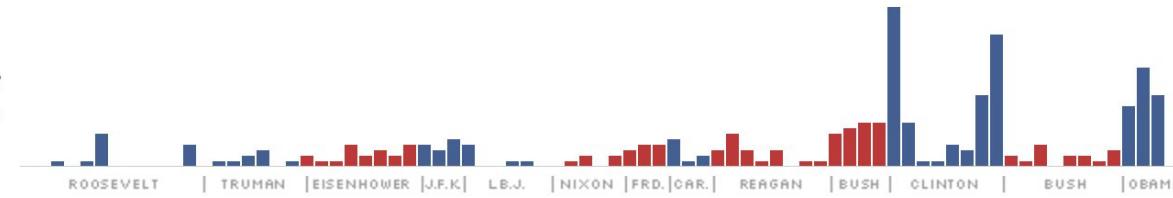
Words included: JOB, JOBS



### ‘invest’

Historically, Democrats use this word more than Republicans, and they mean "public" investment, or new government programs. Bill Clinton used it a lot at the beginning and end of his term, first to propose new programs, and last to take credit for successful ones. Mr. Obama mentioned the word 13 times, proposing new investments in information technology, clean energy and science research.

INVEST, INVESTED, INVESTING, INVESTMENT, INVESTMENTS, INVESTS



**DEMO**

# Concordance

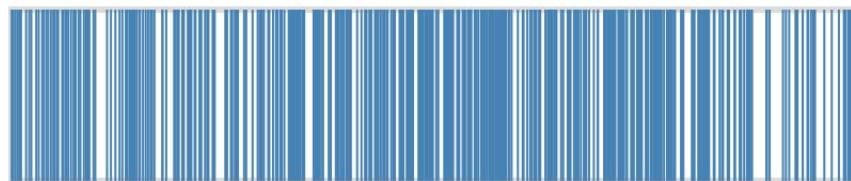
What is the common local context of a term?

Alice was beginning to get very tired of s  
hat is the use of a book , ' thought Alice 'without pictures or conversations ?  
so VERY remarkable in that ; nor did Alice think it so VERY much out of the way  
looked at it , and then hurried on , Alice started to her feet , for it flashed  
hedge . In another moment down went Alice after it , never once considering ho  
ped suddenly down , so suddenly that Alice had not a moment to think about stop  
she fell past it . 'Well ! ' thought Alice to herself , 'after such a fall as t  
own , I think -- ' ( for , you see , Alice had learnt several things of this so  
tude or Longitude I 've got to ? ' ( Alice had no idea what Latitude was , or L  
. There was nothing else to do , so Alice soon began talking again . 'Dinah 'l  
ats eat bats , I wonder ? ' And here Alice began to get rather sleepy , and wen  
dry leaves , and the fall was over . Alice was not a bit hurt , and she jumped  
not a moment to be lost : away went Alice like the wind , and was just in time  
but they were all locked ; and when Alice had been all the way down one side a

# Quick Concordance Plots

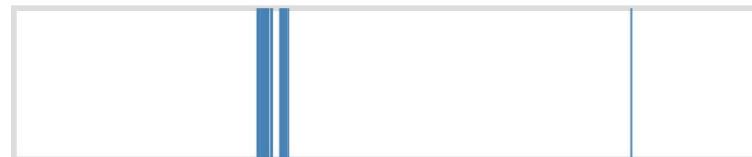
A Concordance plot shows you where in a set of texts a particular search word appears.

Search Word:



that direction waving the other paw lives a march **hare** visit either :

Search Word:

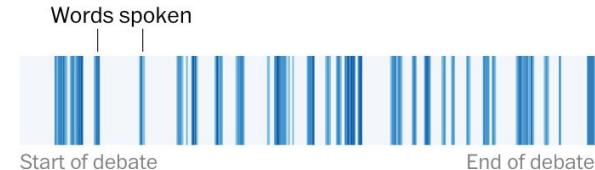


she walked off leaving alice alone with the gryphon **alice** did not q

**DEMO**

# When each candidate spoke

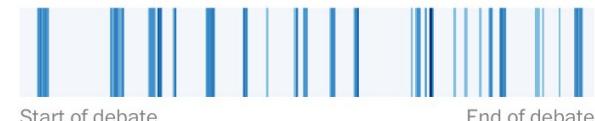
**Hillary Clinton**  
5,397 words



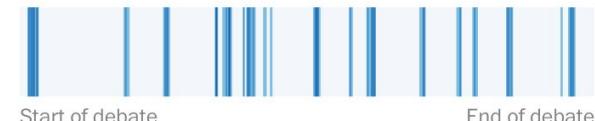
**Bernie Sanders**  
4,561 words



**Martin O'Malley**  
3,232 words



**Jim Webb**  
2,785 words



**Lincoln Chafee**  
1,685 words

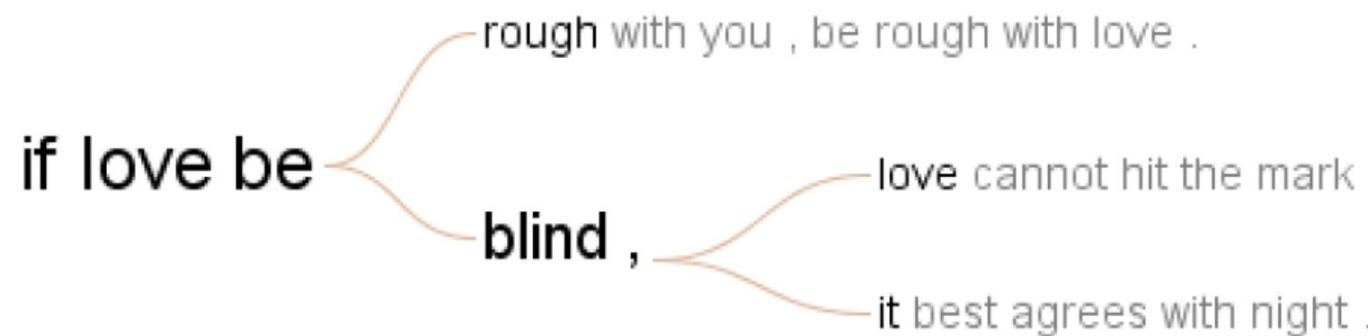


# Word Tree

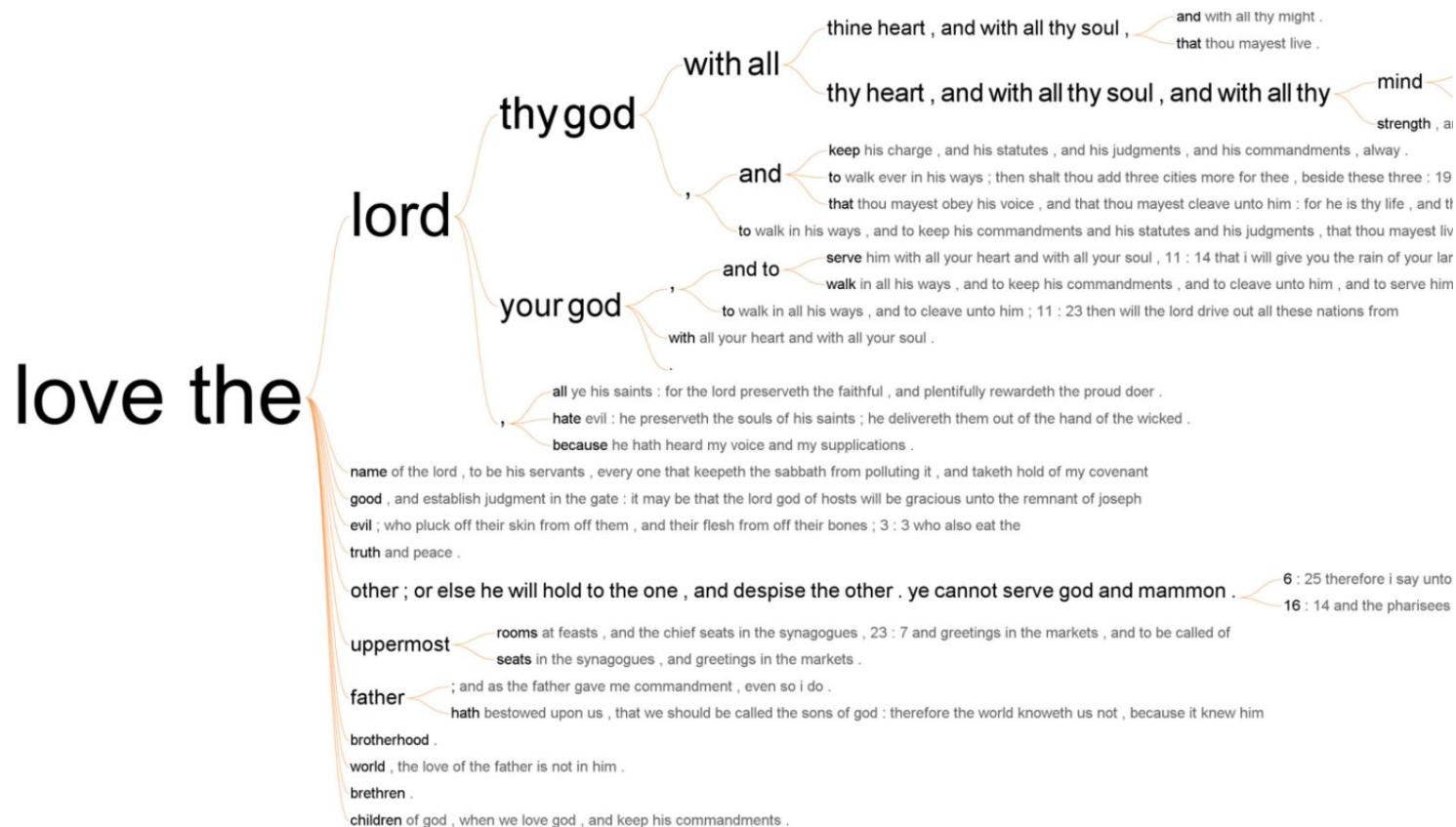
if love be rough with you , be rough with love .

if love be blind , love cannot hit the mark .

if love be blind , it best agrees with night .



# Word Tree [Wattenberg et al.]

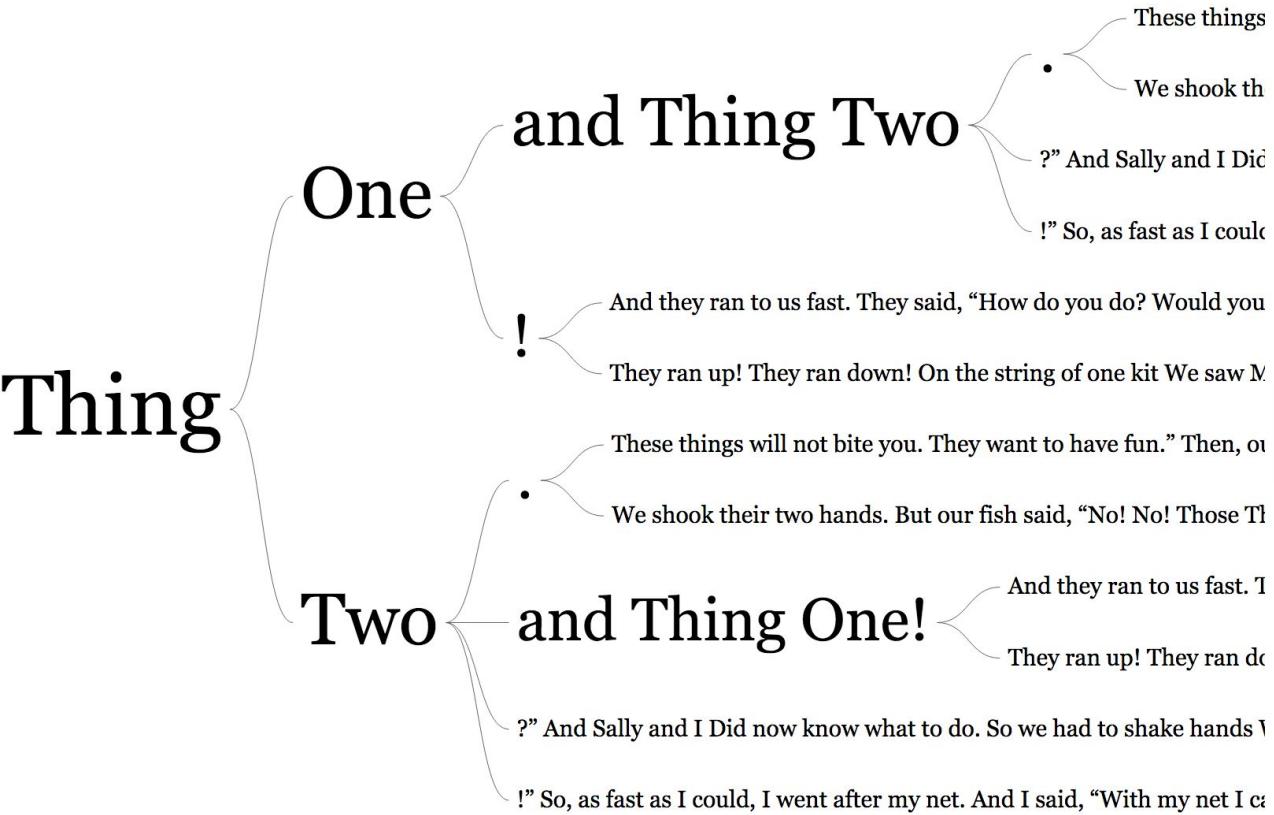


# word tree

Thing

reverse tree  one phrase per line

Shift-click to make that word the root.



hook. You will see something new. Two things. And I call them Thing One and Thing Two. These things will not bite you. They want to have fun." Then, out of the box Came Thing Two and Thing One! And they ran to us fast. They said, "How do you do? Would you like to shake hands With Thing One and Thing Two?" And Sally and I Did now know what to do. So we had to shake hands With Thing One and Thing Two. We shook their two hands. But our fish said, "No! No! Those Things should not be In this house! Make them go! "They should not be here When your mother is not! Put them out! Put them out!" Said the fish in the pot. "Have no fear, little fish," Said the Cat in the Hat. "These things are good Things." And he gave them a pat. "They are tame. Oh, so tame! They have

**DEMO**

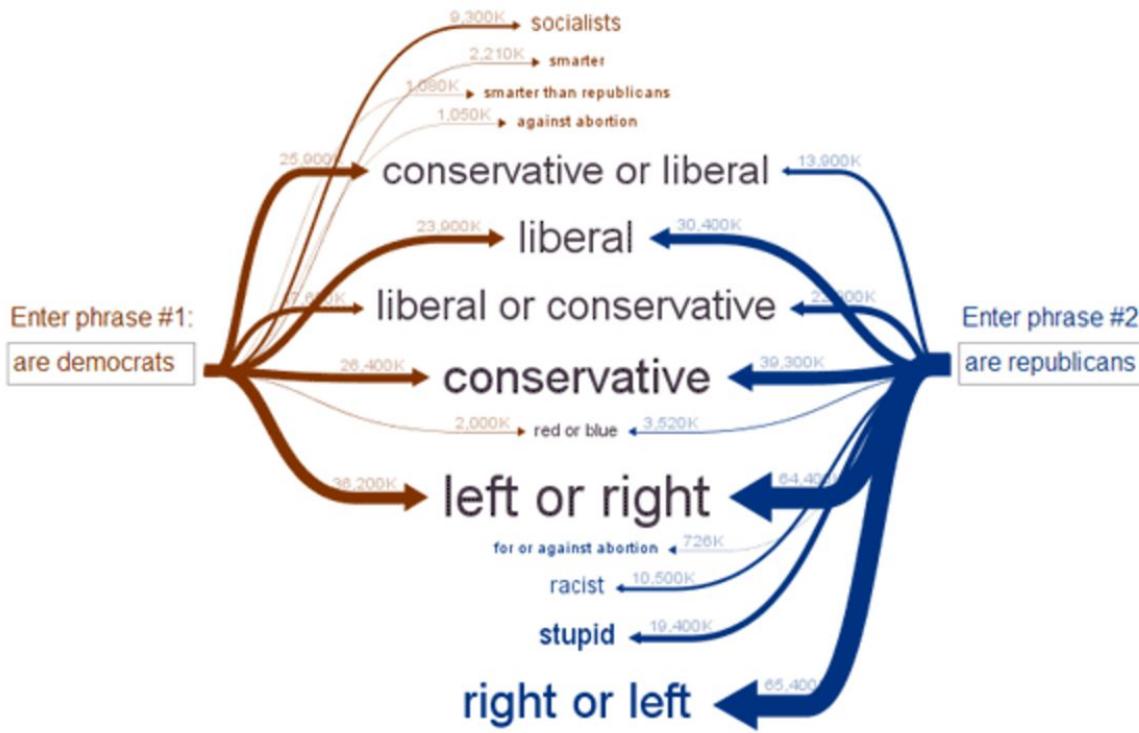
# Web Seer



**Is there a way to visualize people's innermost thoughts?**

Google Suggest lets you see what others are asking when they search the web. From the existential to the mundane, the questions form a portrait of human curiosity

# Web Seer



**DEMO**

# Phrase Nets [van Ham et al.]

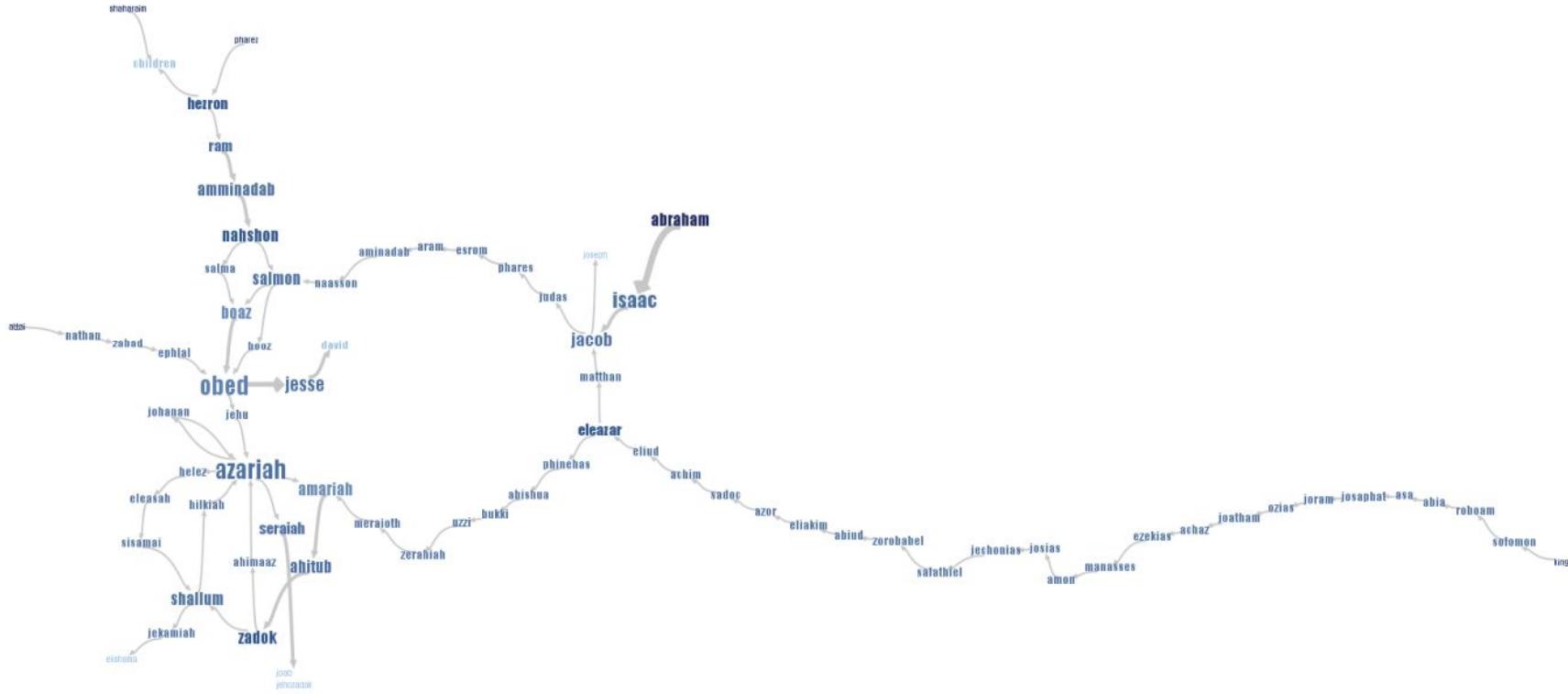
Look for specific linking patterns in the text:

- ‘A and B’, ‘A at B’, ‘A of B’, etc
- Could be output of regexp or parser.

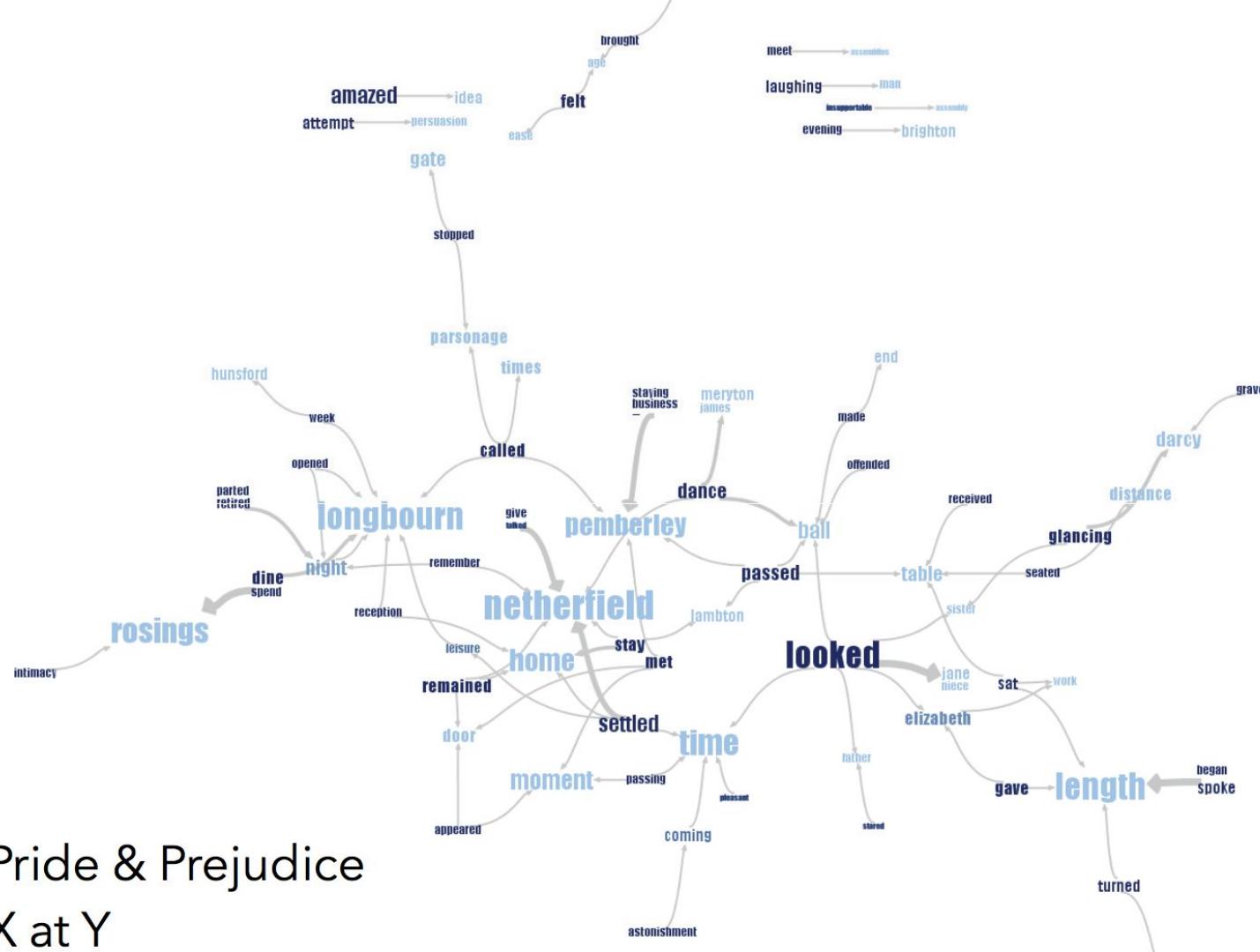
Visualize patterns in a node-link view

- Occurrences -> Node size
- Pattern position -> Edge direction

# Phrase Nets



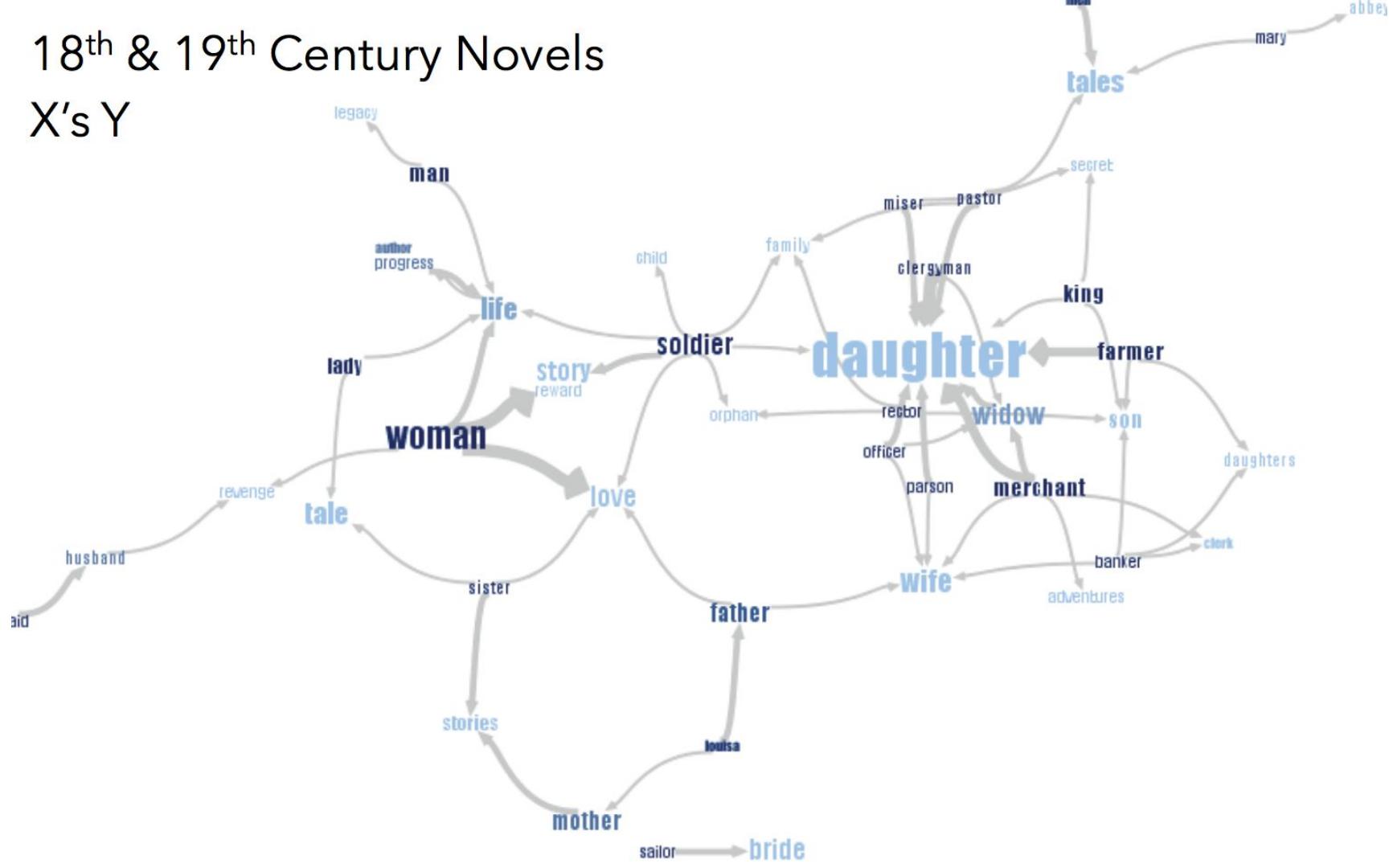
Scanning the bible for textual matches to the pattern 'X begat Y' reveals a network of family relations



Pride & Prejudice  
X at Y

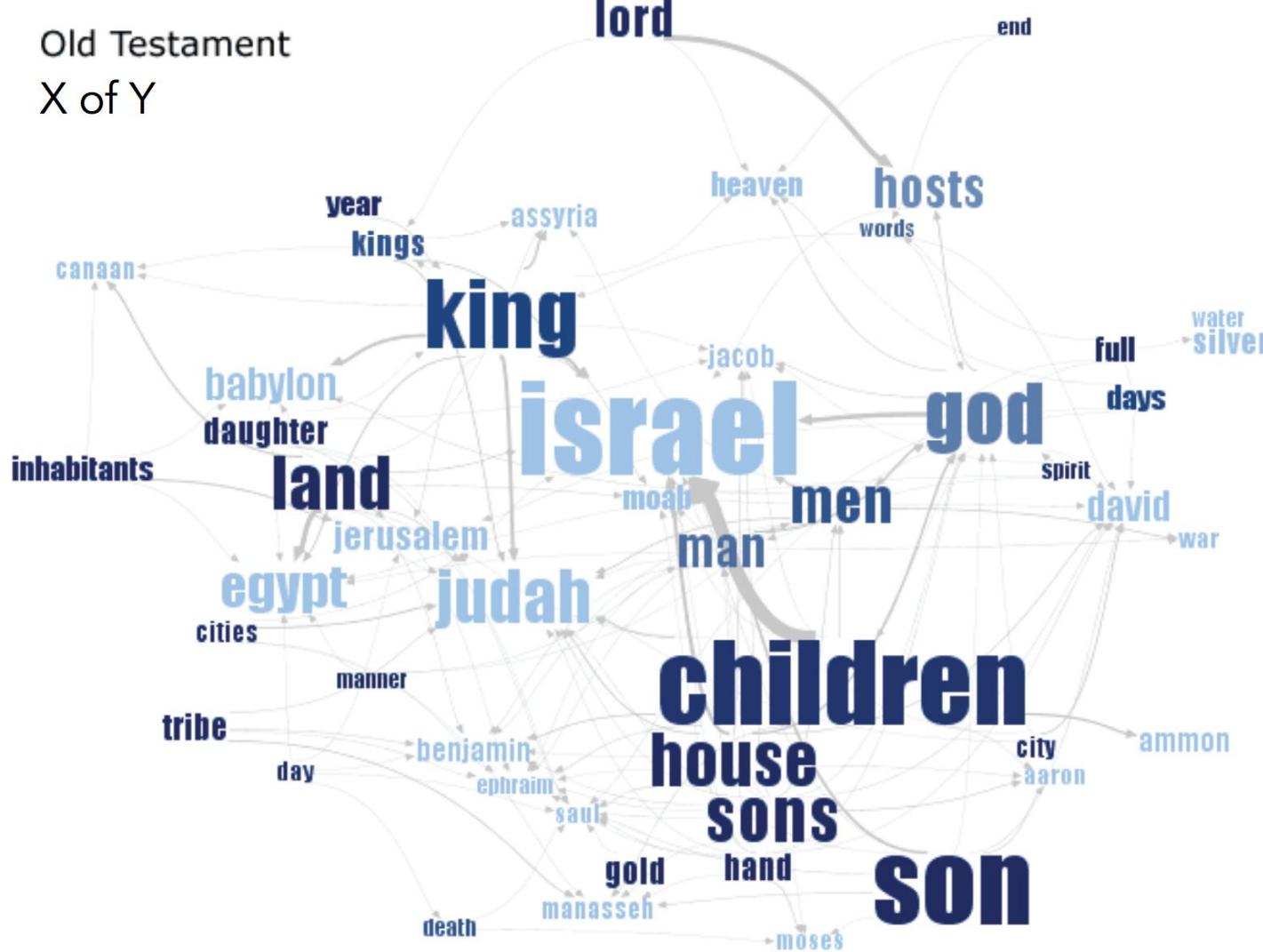
## 18<sup>th</sup> & 19<sup>th</sup> Century Novels

X's Y



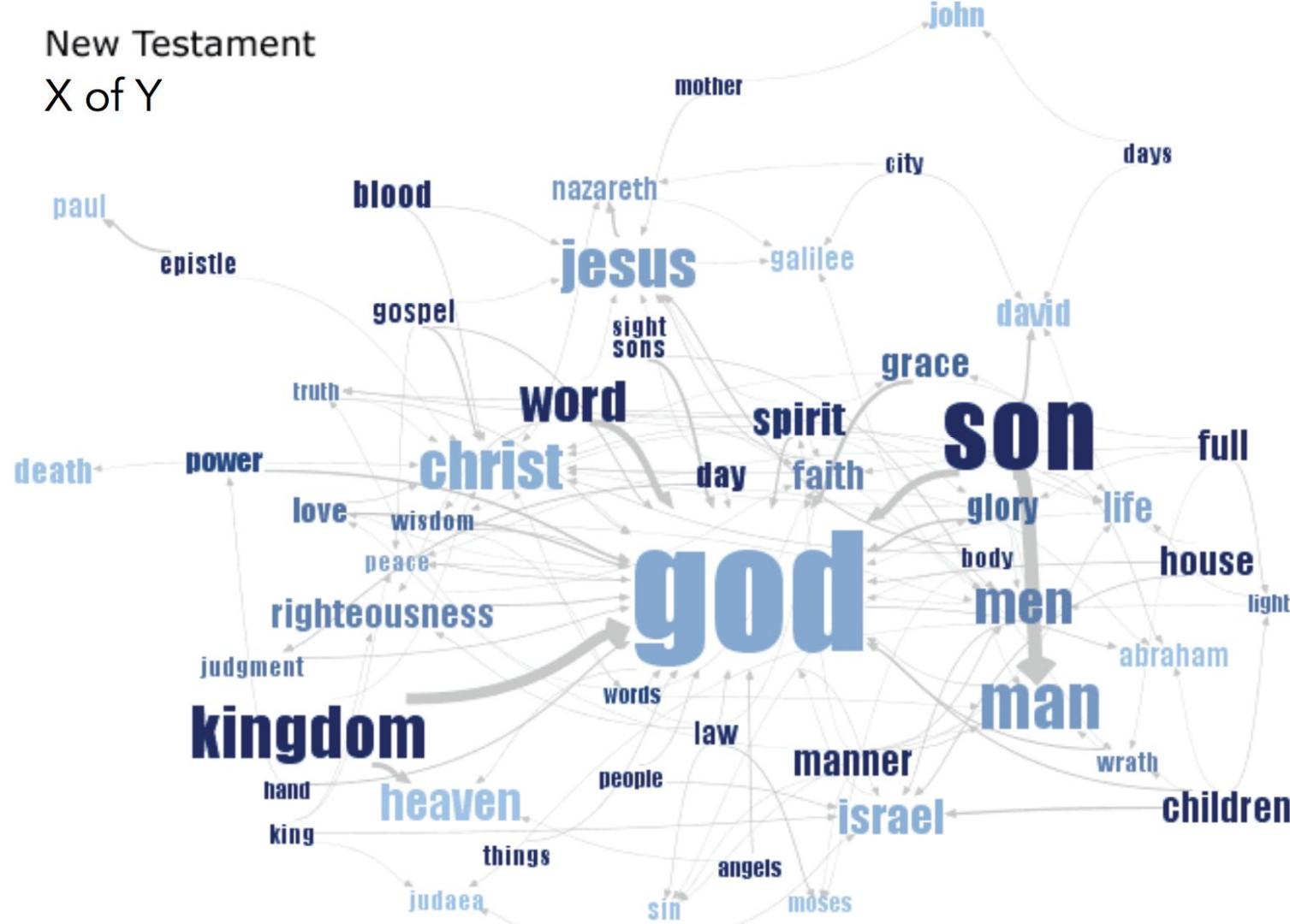
Old Testament

X of Y



# New Testament

X of Y



# Document Content

## Understand Your Analysis Task

- Visually: Word position, browsing, brush & link
- Semantically: Word sequence, hierarchy, clustering
- Both: Spatial layout reflects semantic relationships

# Evolving Documents

# Visualizing Revision History

How to depict contributions over time?

Example: Wikipedia history log

## Chocolate

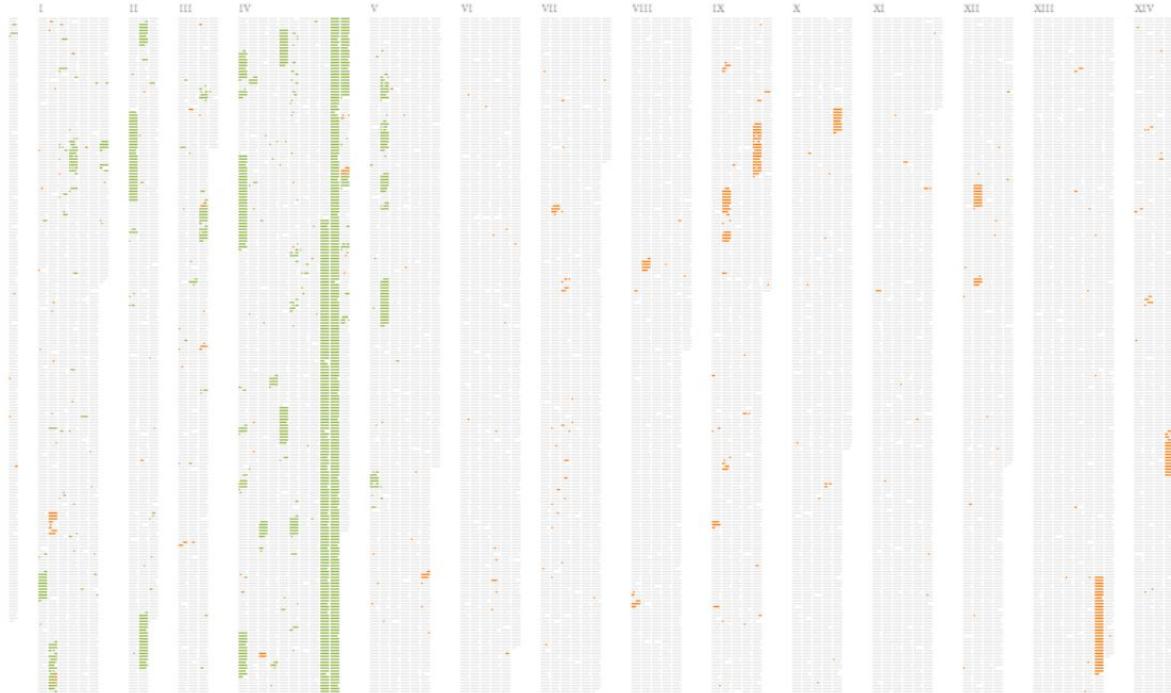
Revision history

Legend: (cur) = difference with current version, (last) = difference with preceding version, M = minor edit

- (cur) (last) . . 12:01, 20 Aug 2003 . . [Dysprosia](#) (*neaten to do, rearrange see also*)
- (cur) (last) . . 11:59, 20 Aug 2003 . . [Patrick](#)
- (cur) (last) . . 11:52, 20 Aug 2003 . . [81.203.98.109](#)
- (cur) (last) . . M 18:36, 6 Aug 2003 . . [Manika](#) (*corrected spelling*)
- (cur) (last) . . 18:32, 6 Aug 2003 . . [Daniel Quinlan](#) (*removing obscure heraldry information, belongs on [[heraldry]] if anywhere*)
- (cur) (last) . . 15:21, 6 Aug 2003 . . [Rmhermen](#)
- (cur) (last) . . 15:08, 6 Aug 2003 . . [Cyp](#) (*Chocolate often has odd shapes.*)
- (cur) (last) . . 19:14, 3 Aug 2003 . . [Daniel C. Boyer](#) ("chocolate" as shade of gules in heraldry)
- (cur) (last) . . M 02:00, 30 Jul 2003 . . [Evercat](#) (*fmt*)

# Animated Traces [Ben fry]

ON THE ORIGIN OF SPECIES *The Preservation of Favoured Traces*



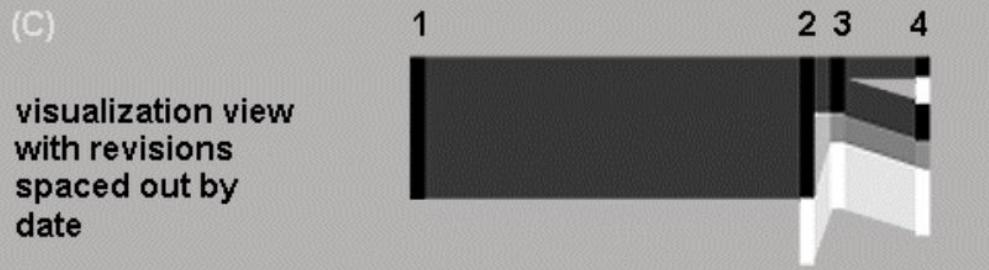
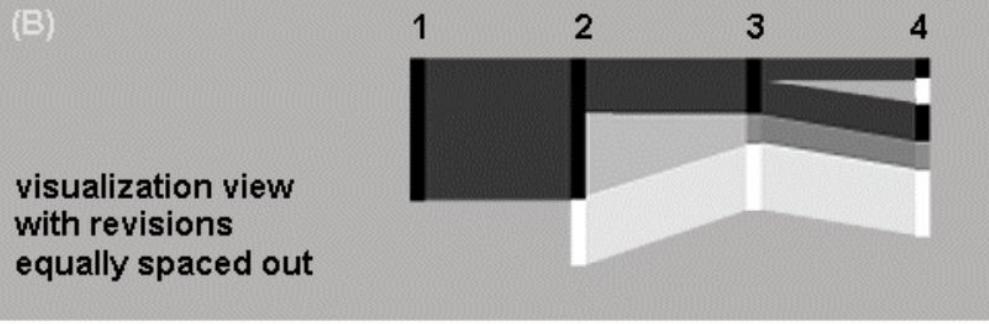
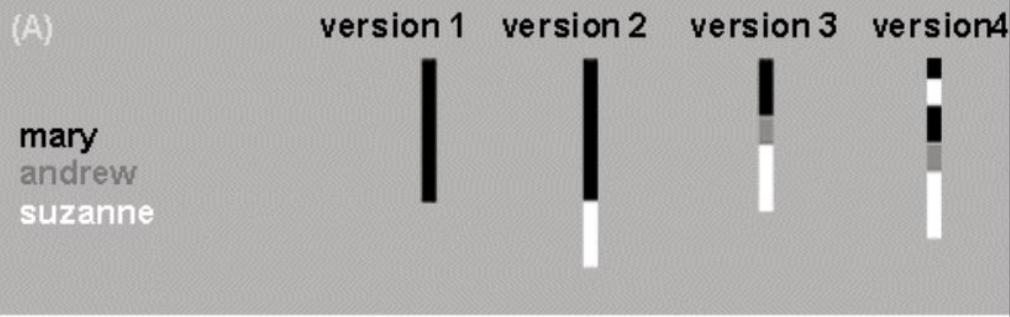
First Edition (1859)    Second Edition (1860)    Third Edition (1861)    Fourth Edition (1866)    Fifth Edition (1869)    Sixth Edition (1872)

<https://fathom.info/traces/>

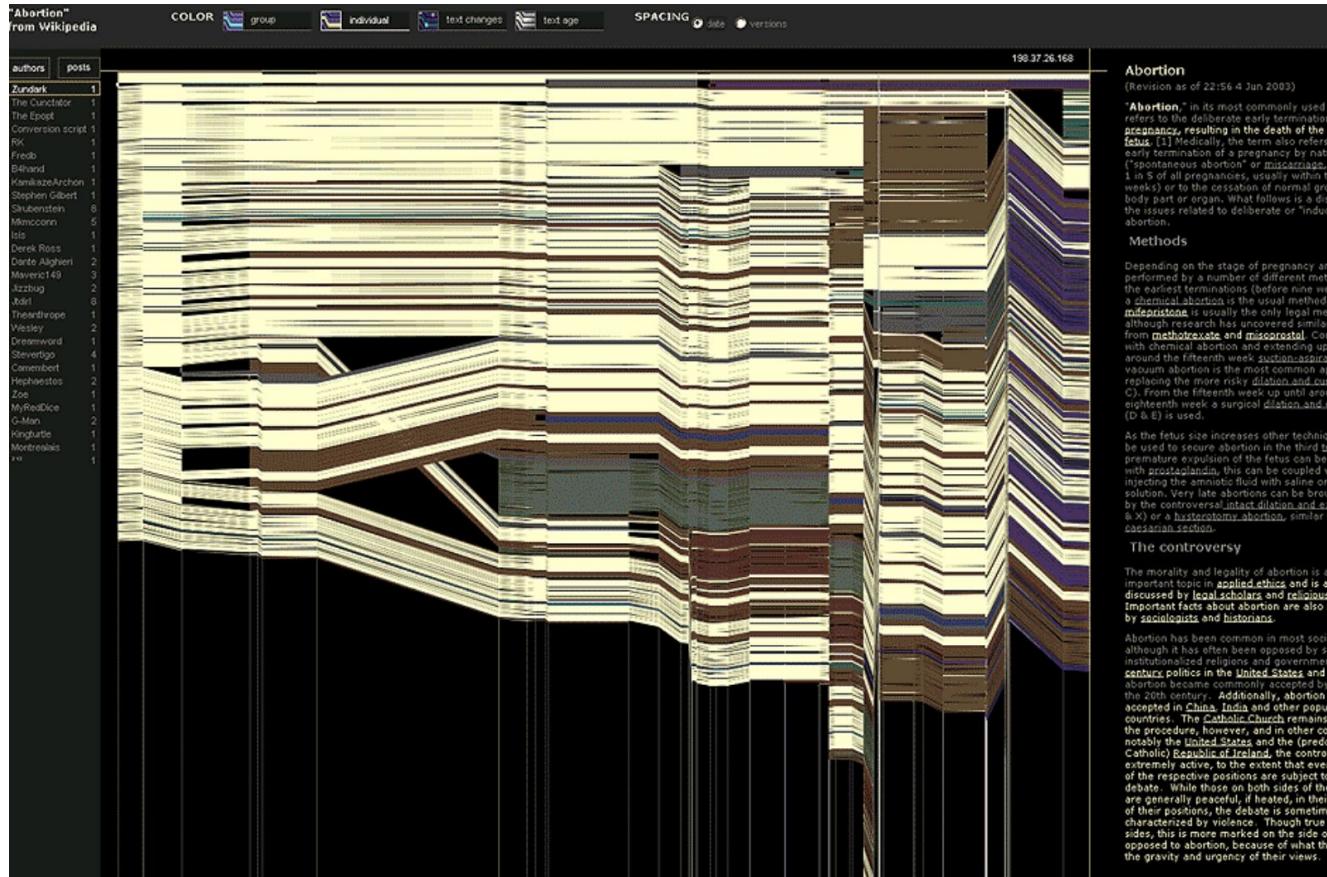
**DEMO**

# History Flow

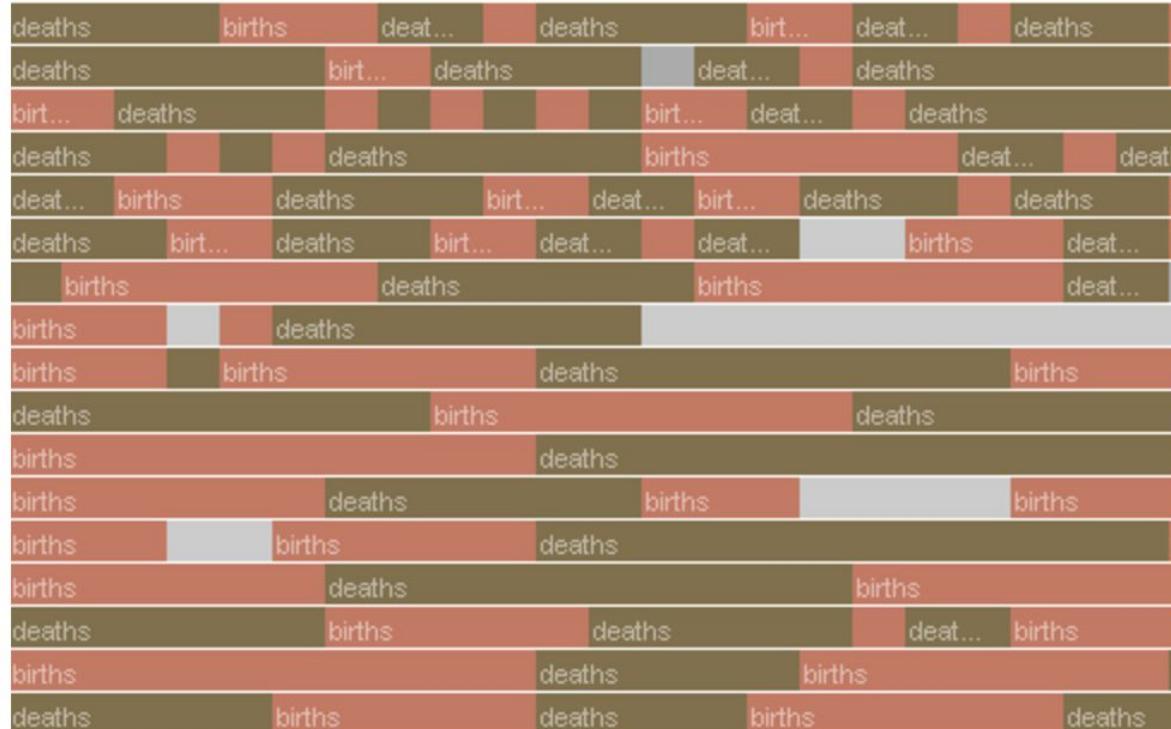
[Viegas et al.]



# Wikipedia History Flow



# Chromogram: how participants in Wikipedia allocate their time



A Chromogram of a user whose edits are mostly about adding births and deaths occurrences to year articles on Wikipedia.

# Creating a Chromogram



a

Time	Page	Comment
May 21, 9:32 am	Sphere	Add cite
May 22, 10:56 am	Sphere	New Intro
May 22, 1:23 pm	sphere.png	Copyright
May 22, 2:54 pm	Helix	Spell check
May 22, 3:00 pm	Mathematics	Revert
May 24, 11:21 am	FIELDS Medal	2006 data
May 24, 11:25 am	Talk:Fractal	List of proposed changes
May 24, 11:27 am	Talk:Fractal	List of proposed changes
May 25, 10:13 am	Sphere	Fix tpyo
May 25, 10:23 am	mercator.png	Copyright

b Chromogram of comments:  
Timeline View



The first three letters of a string determine color in a Chromogram. The first letter determines the hue; the second letter the saturation, and the third the brightness.

# Chromogram: how participants in Wikipedia allocate their time



A Chromogram of a user focused on naval history.

# Chromogram: how participants in Wikipedia allocate their time



Comment chromogram of about 2,900 copy edits. Green items are changes from “it’s” to “its”

# **Conversations**

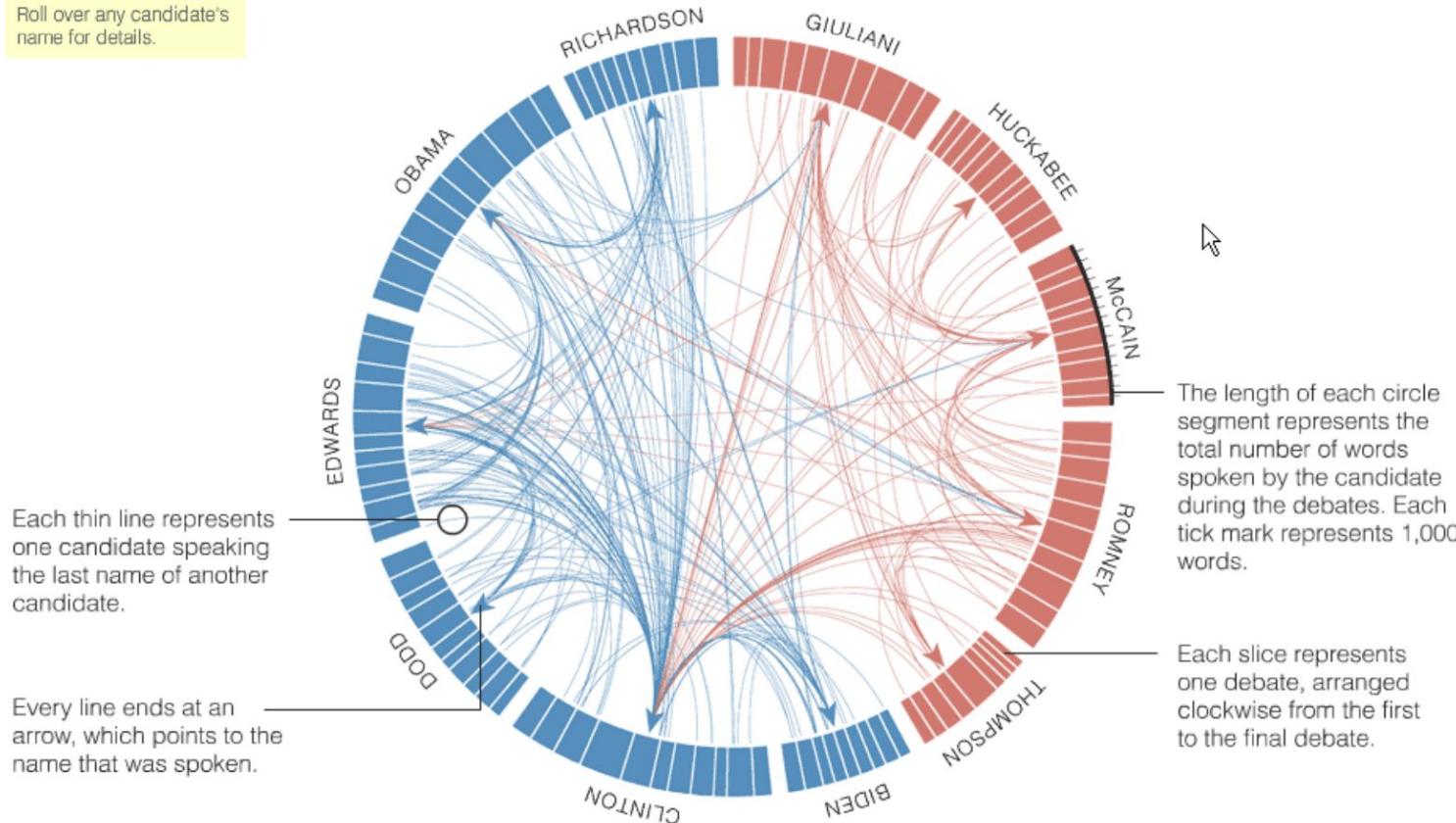
# Visualizing Conversation

- Many dimensions to consider:
  - Who (senders, receivers)
  - What (the content of communication)
  - When (temporal patterns)
- Interesting cross-products:
  - What x When -> Topic “Zeitgeist”
  - Who x Who -> Social network
  - Who x Who x What x When -> Information flow

# Naming Names

Names used by major presidential candidates in the series of Democratic and Republican debates leading up to the Iowa caucuses.

Roll over any candidate's name for details.

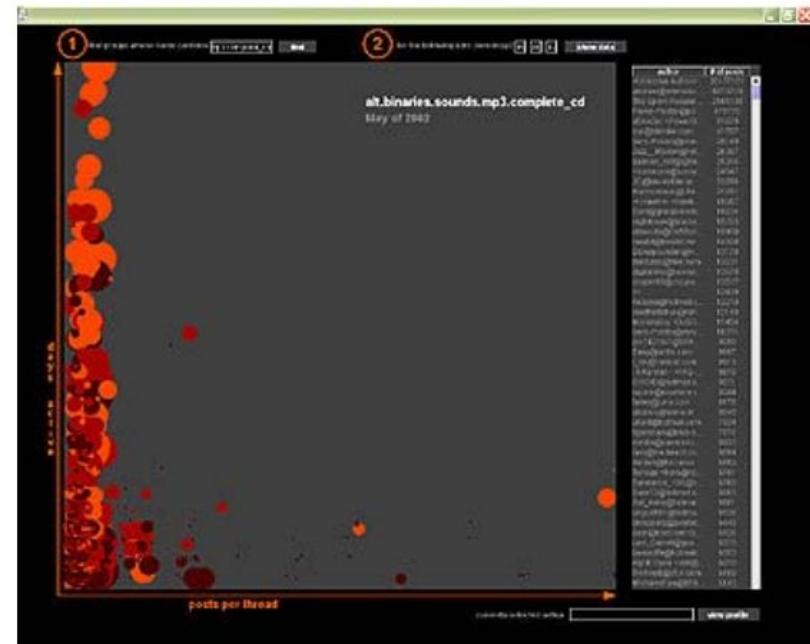


# Usenet Visualization [Viegas & Smith]

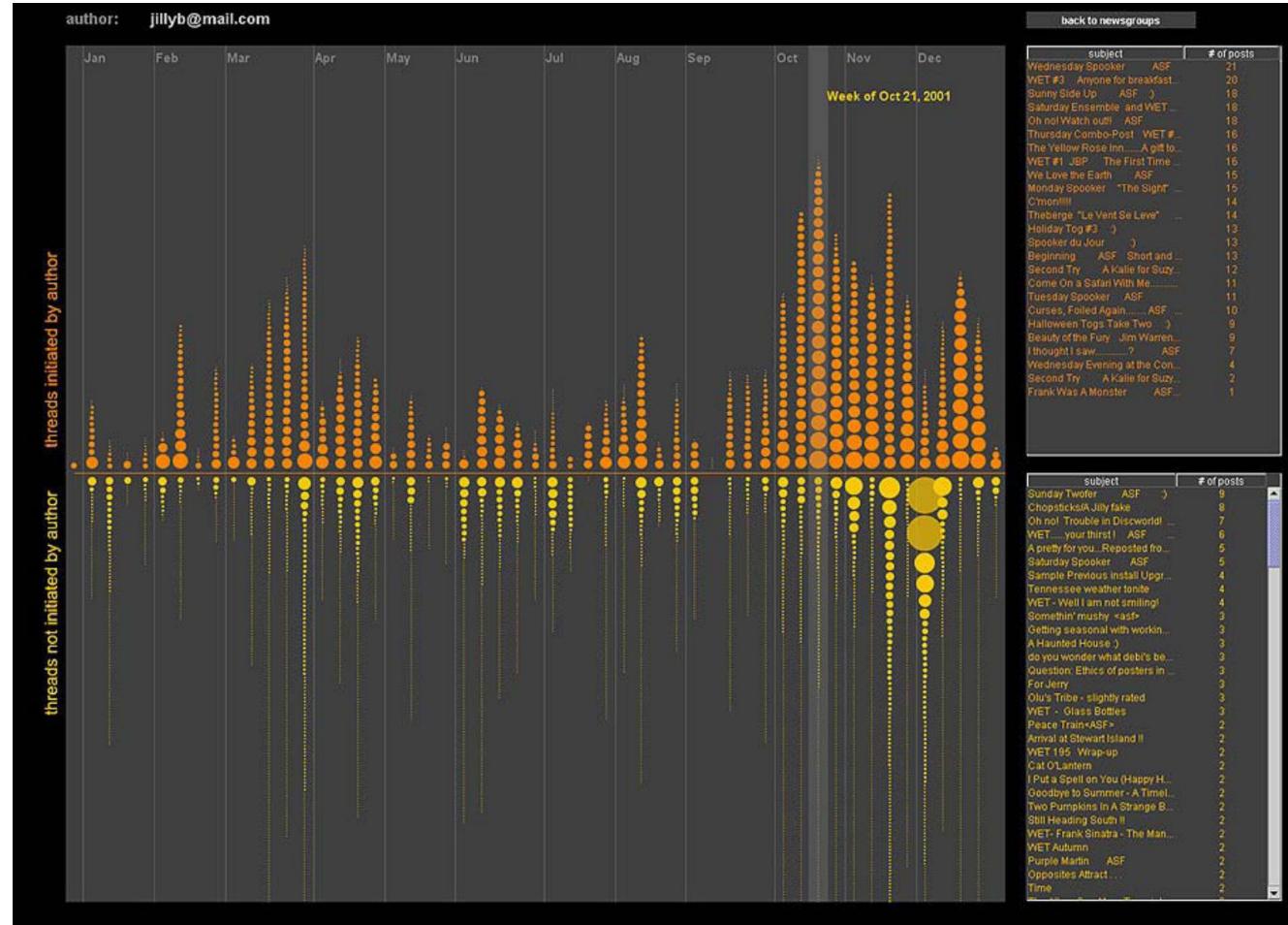
Show correspondence patterns in text forums

x: average post per thread

y: number of active days



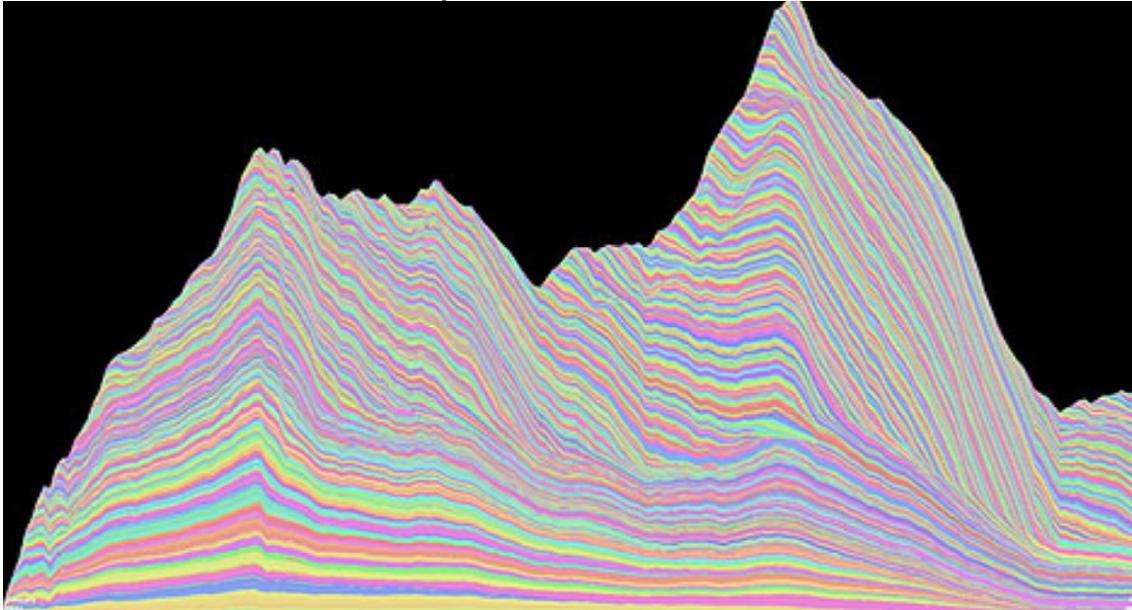
# Metrics of authors' activity



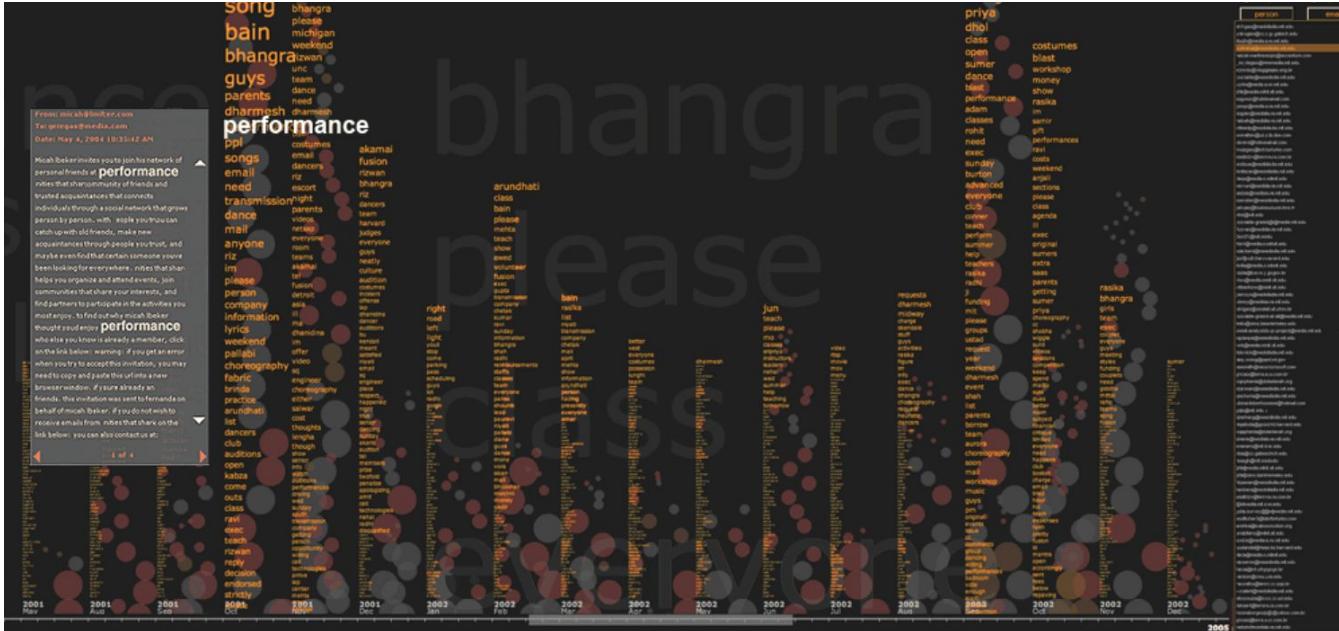
# Email Mountain [Viegas]

Conversation by person over time (who x when).

Mountain visualizes a person's email archive in terms of all the people with whom this person has been in touch over the years.



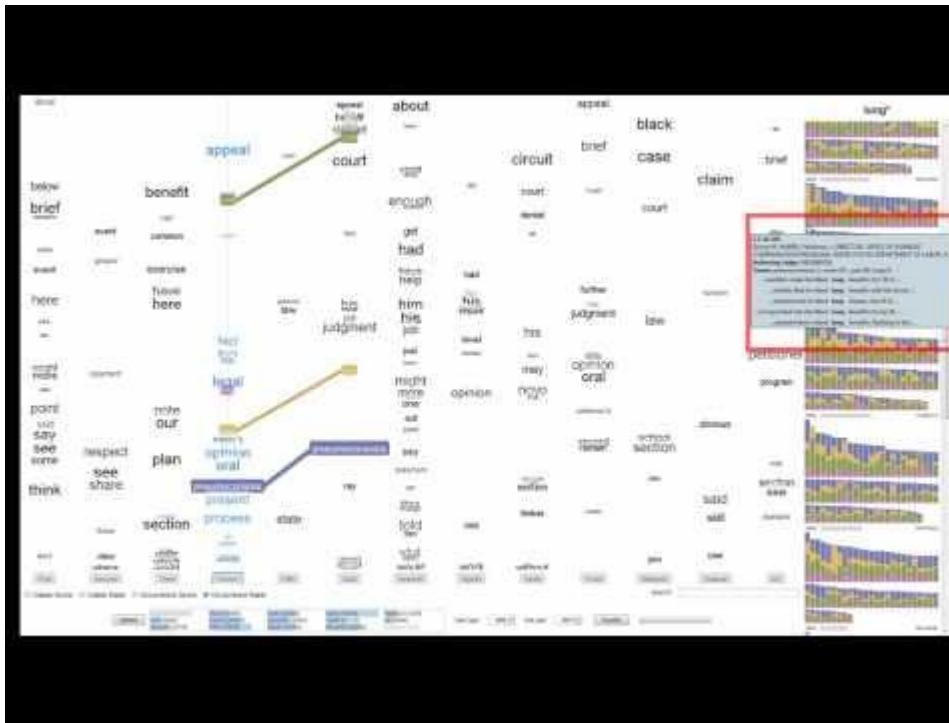
# The mail [Viegas]



# One person over time, TF.IDF weighted terms

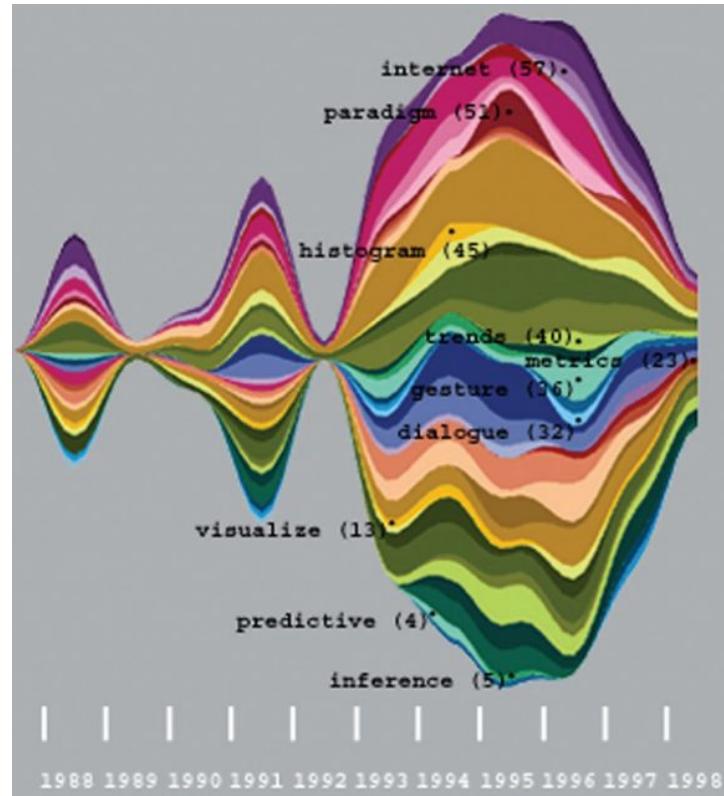
# Document Collections

# Parallel tag clouds

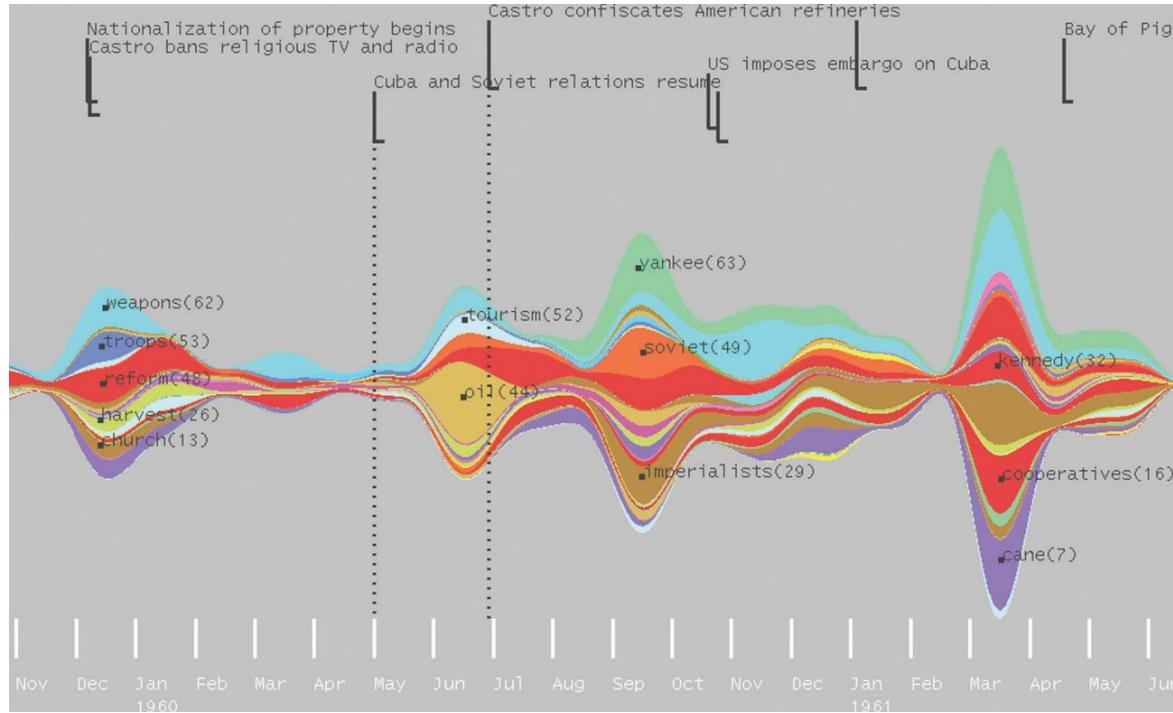


# Theme River [Havre et al.]

Shows thematic changes over time in a collection of patents from one company



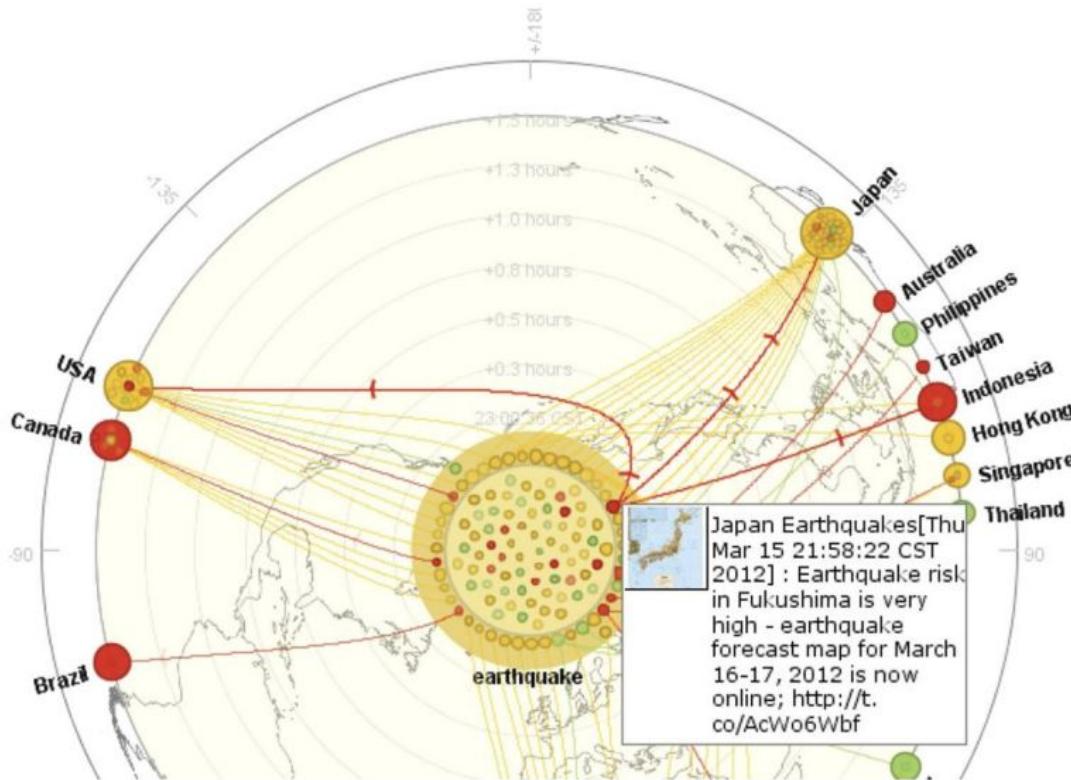
# Theme River [Havre et al.]



Collection of Fidel Castro speeches, interviews and articles

**Other**

# Whisper: information diffusion on Twitter

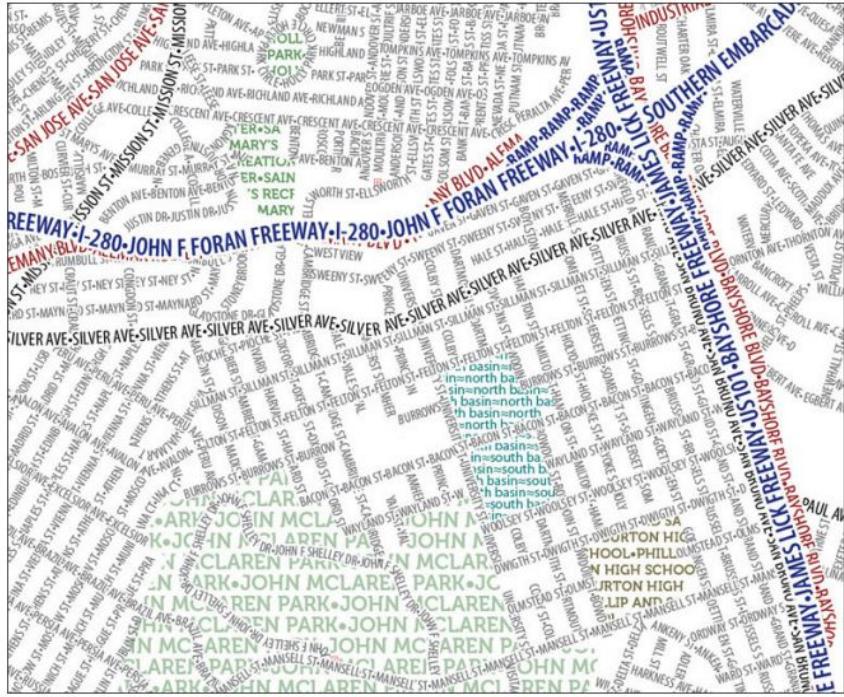


<https://www.youtube.com/watch?v=ou8L0MzGvOU>

-  RT @kasakoworld: 千葉東方沖地震  
は3/3~3/10まで一度もなかつた。3/14
-  Christofer  
Thu Mar 15 23:32:10 HKT 2012  
RT @WhatTheFFacts: The Fukushima  
earthquake sped the Earth's rotation
-  ノーリツチテリアのあたまる  
Thu Mar 15 23:29:50 HKT 2012  
RT @kasakoworld: 千葉東方沖地震  
は3/3~3/10まで一度もなかつた。3/14
-  That Girl♥  
Thu Mar 15 23:29:38 HKT 2012  
RT @FreddyAmazin: \*Phone vibrates at  
home\* barely hear it \*Phone vibrates
-  • ω •  
Thu Mar 15 23:27:53 HKT 2012  
RT @begining\_of\_end: 千葉県東方沖震  
源の多さ
-  andy perkins  
Thu Mar 15 23:27:26 HKT 2012  
RT @wschamberlain: On stage mid set  
in Tokyo a 6.5 scale earthquake
-  Kathleen Morgan  
Thu Mar 15 23:25:03 HKT 2012  
RT @EpicTweets\_: \*Phone vibrates at  
home\* barely hear it \*Phone vibrates

**DEMO**

# Typographic Maps



Typographic maps merge text and spatial data into a visual representation where text alone forms the graphical features.

# Thema-Typographic Maps

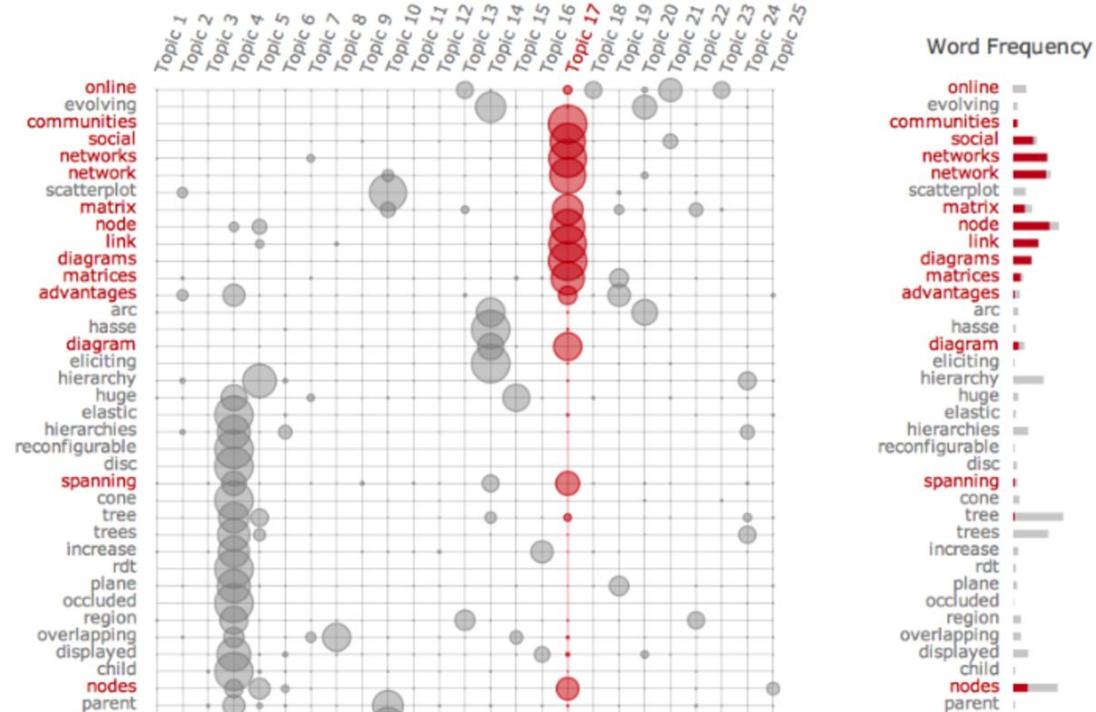


Statistical variable  
being visualized is  
crime rate.  
Larger text means  
more crime.

# Visualization of Textual Topic Models

Displays term-topic distributions for an LDA topic model.

A bar chart shows the marginal probability of each term.



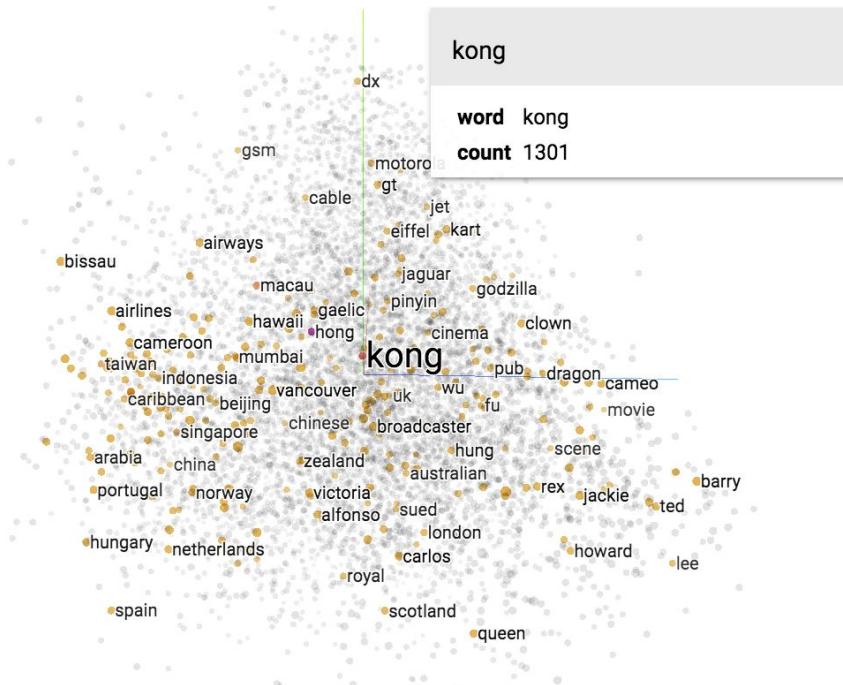
**Word embedding** is the collective name for a set of language modeling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers.

banana



# Embedding Projector:

## Interactive Visualization and Interpretation of Embeddings



<http://projector.tensorflow.org/>

**DEMO**

# References

- <http://courses.cs.washington.edu/courses/cse512/16sp/lectures/CSE512-Text.pdf>
- <http://shixiali.com/publications/InfovisSurvey/paper.pdf>