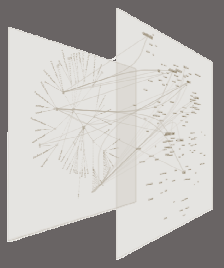
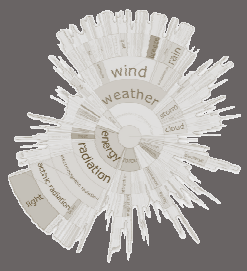
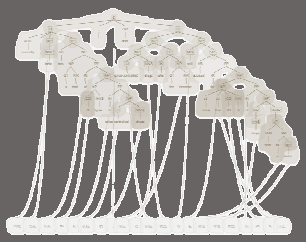





Interactive Visualization for Computational Linguistics

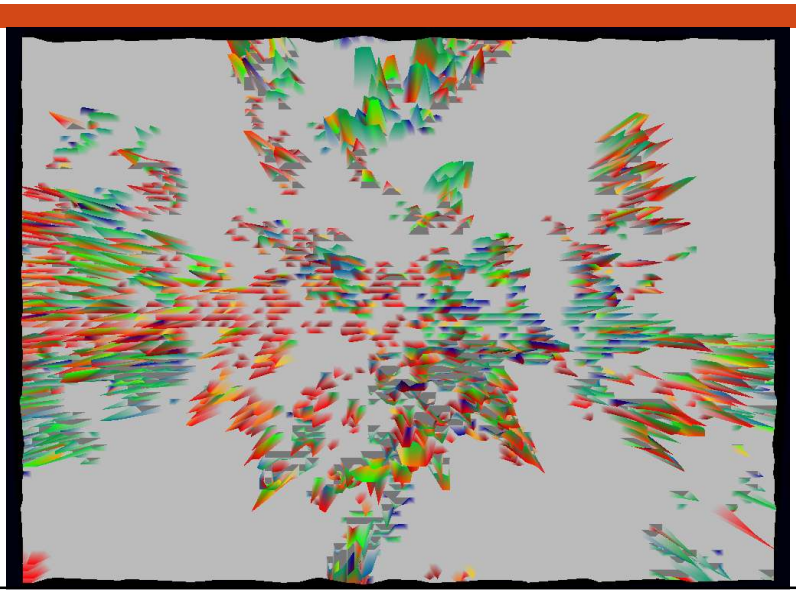
ESSLII 2009

2

Review of Linguistic Visualizations

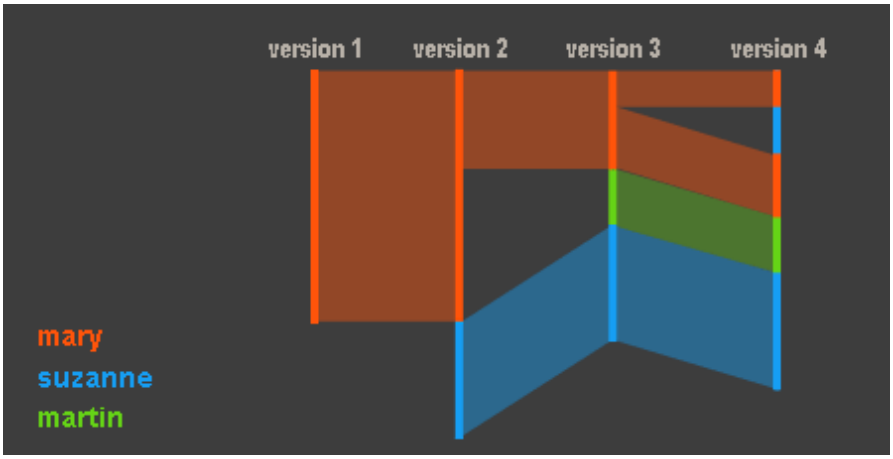
Non-textual linguistic data

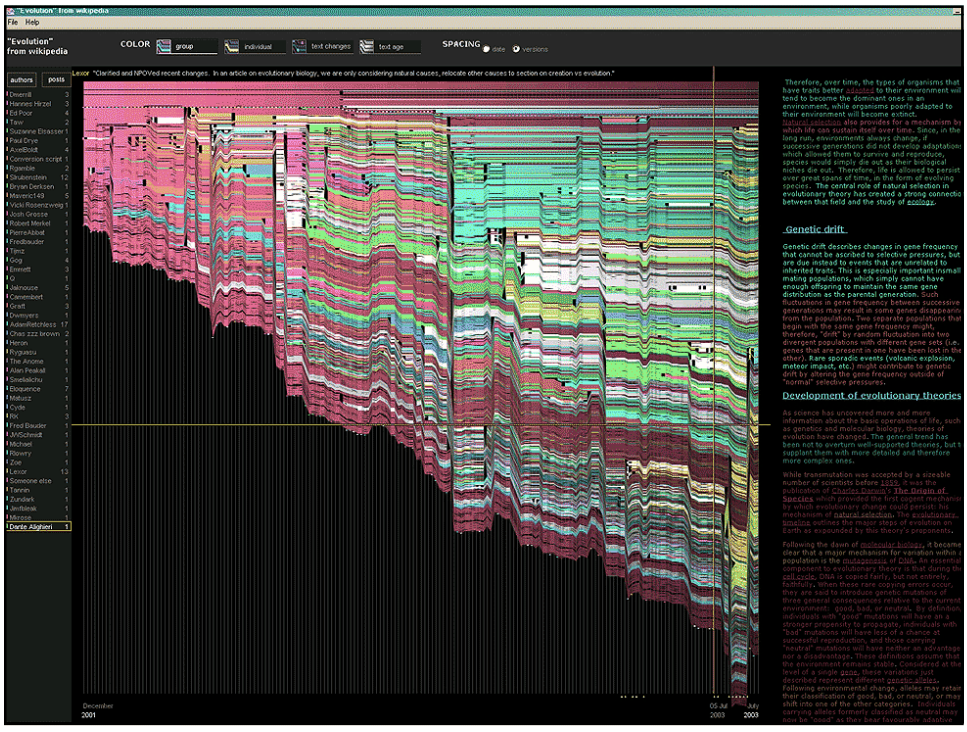
Shape Of Conversation

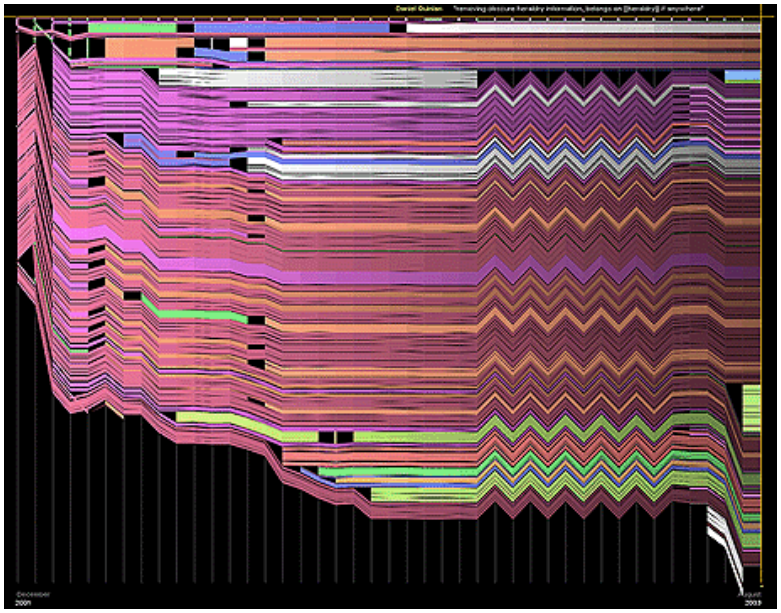


Tse, Simon 2003

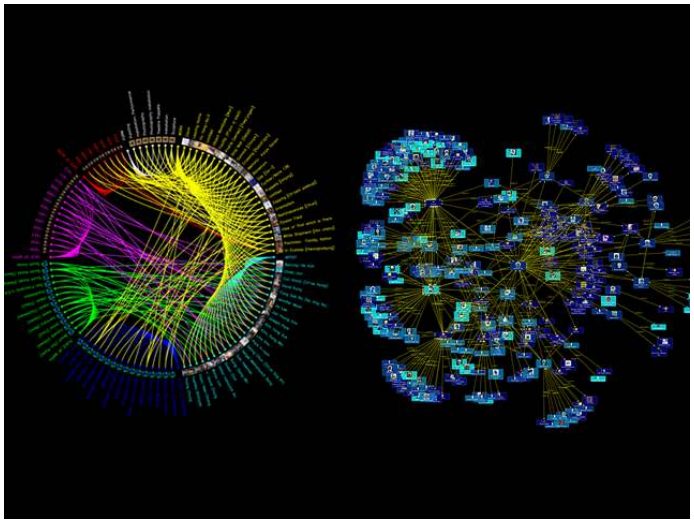
Linguistic and Non-linguistic Data Together

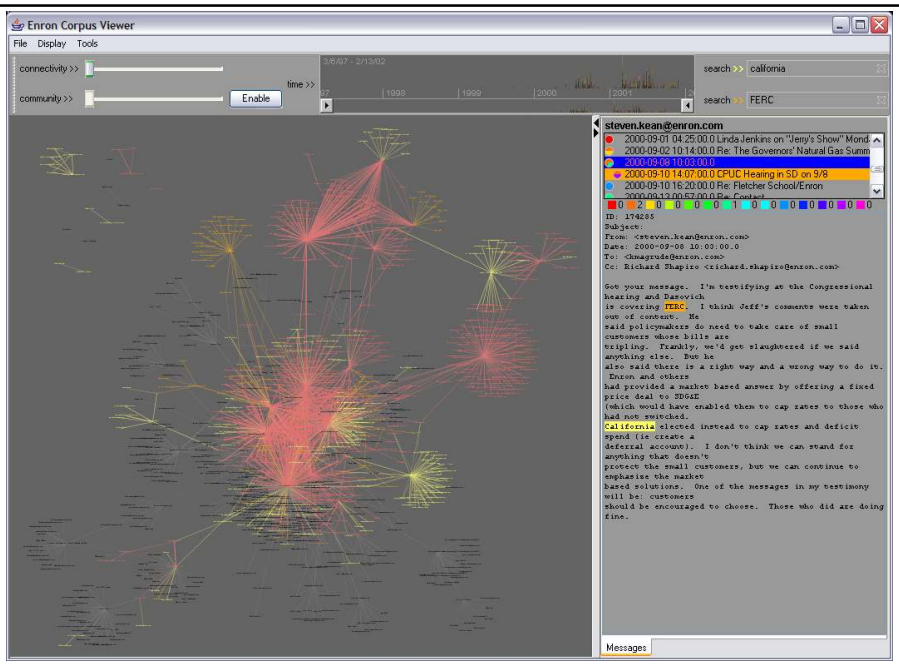

version 1
version 2
version 3
version 4
mary
suzanne
martin


"Evolution" from wikipedia
File Help
"Evolution"
from wikipedia
COLOR group individual text changes text age
SPACING date versions
authors posts
Levor "Clarified and NPOVed recent changes. In an article on evolutionary biology, we are only considering natural causes, relocate other causes to section on creation vs evolution."
Therefore, over time, the types of organisms that have traits better adapted to their environment will tend to become the dominant ones in an environment, while organisms poorly adapted to their environment will become extinct.
Natural selection also provides for a mechanism by which life can sustain itself over time. Since, in the long run, environments always change, if successive generations did not develop adaptations which allowed them to survive and reproduce, species would simply die out as their biological niches die out. Therefore, life is allowed to persist over great spans of time, in the forms of evolving species. The central role of natural selection in evolutionary theory has created a strong connection between that field and the study of ecology.
Genetic drift
Genetic drift describes changes in gene frequency that cannot be ascribed to selective pressures, but are due instead to events that are unrelated to inherited traits. This is especially important in small mating populations, which simply cannot have enough offspring to maintain the same gene distribution as the parental generation. Such fluctuations in gene frequency between successive generations may result in some genes disappearing from the population. Two separate populations that begin with the same gene frequency might, therefore, "drift" by random fluctuation into two divergent populations with different gene sets (i.e. genes that are present in one have been lost in the other). Rare sporadic events (volcanic explosion, meteor impact, etc.) might contribute to genetic drift by altering the gene frequency outside of "normal" selective pressures.
Development of evolutionary theories
As science has uncovered more and more information about the basic operations of life, such as genetics and molecular biology, theories of evolution have changed. The general trend has been not to overturn well-supported theories, but to supplant them with more detailed and therefore more complex ones.
While transmutation was accepted by a sizeable number of scientists before 1859, it was the publication of Charles Darwin's The Origin of Species which provided the first cogent mechanism by which evolutionary change could persist: his mechanism of natural selection. The evolutionary timeline outlines the major steps of evolution on Earth as expounded by this theory's proponents.
Following the dawn of molecular biology, it became clear that a major mechanism for variation within a population is the mutagenesis of DNA. An essential component to evolutionary theory is that during cell cycle, DNA is copied fairly, but not entirely, faithfully. When these rare copying errors occur, they are said to introduce genetic mutations of three general consequences relative to the current environment: good, bad, or neutral. By definition, individuals with "good" mutations will have an a stronger propensity to propagate, individuals with "bad" mutations will have less of a chance at successful reproduction, and those carrying "neutral" mutations will have neither an advantage nor a disadvantage. These definitions assume that the environment remains stable. Considered at the level of a single gene, these variations just described represent different genetic alleles. Following environmental change, alleles may retain their classification of good, bad, or neutral, or may shift into one of the other categories. Individuals carrying alleles formerly classified as neutral may now be "good" as they bear favourable adaptive
December 2001
05 Jul 2003
2003



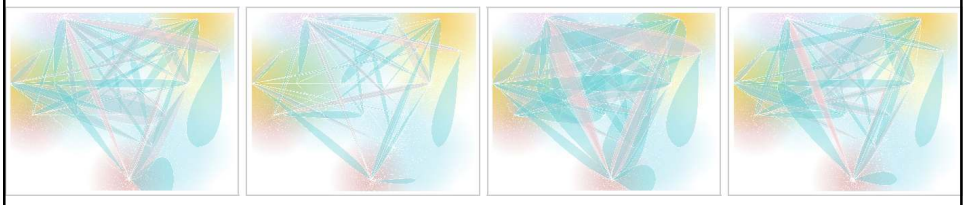
Many Visualizations on Social Relationships…





Enriching digital communication

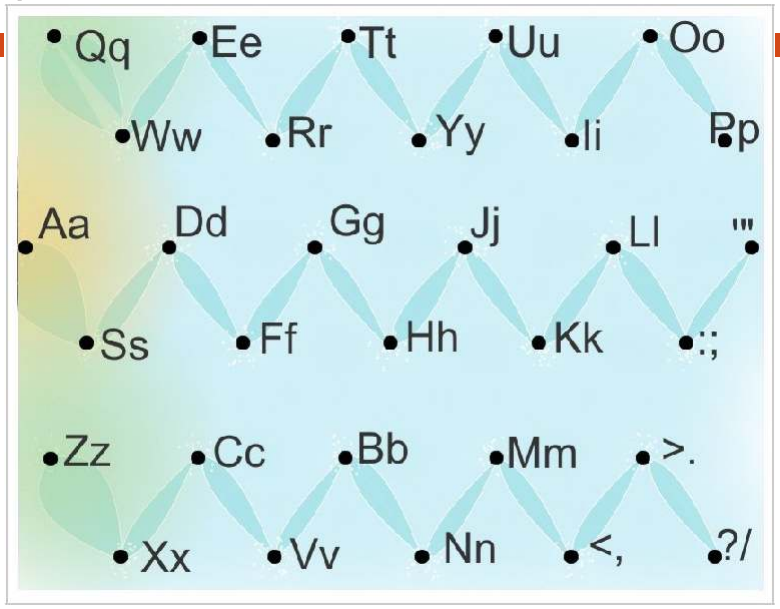
Keystrokes



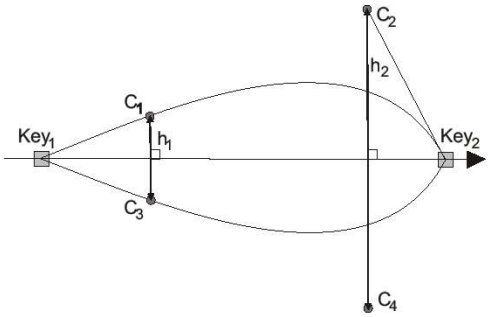
Petra Neumann, Annie Tat, Torre Zuk, Sheelagh Carpendale

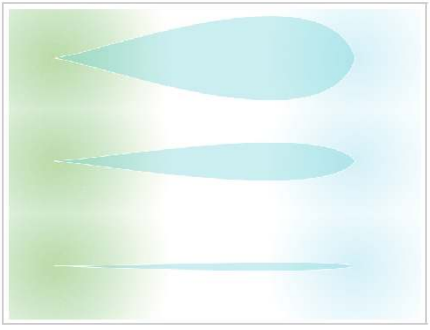
Department of Computer Science, University of Calgary,

Keystrokes



Keystrokes



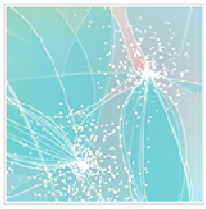


Message Patterns

Erasing

Key Combinations

Individual Keys



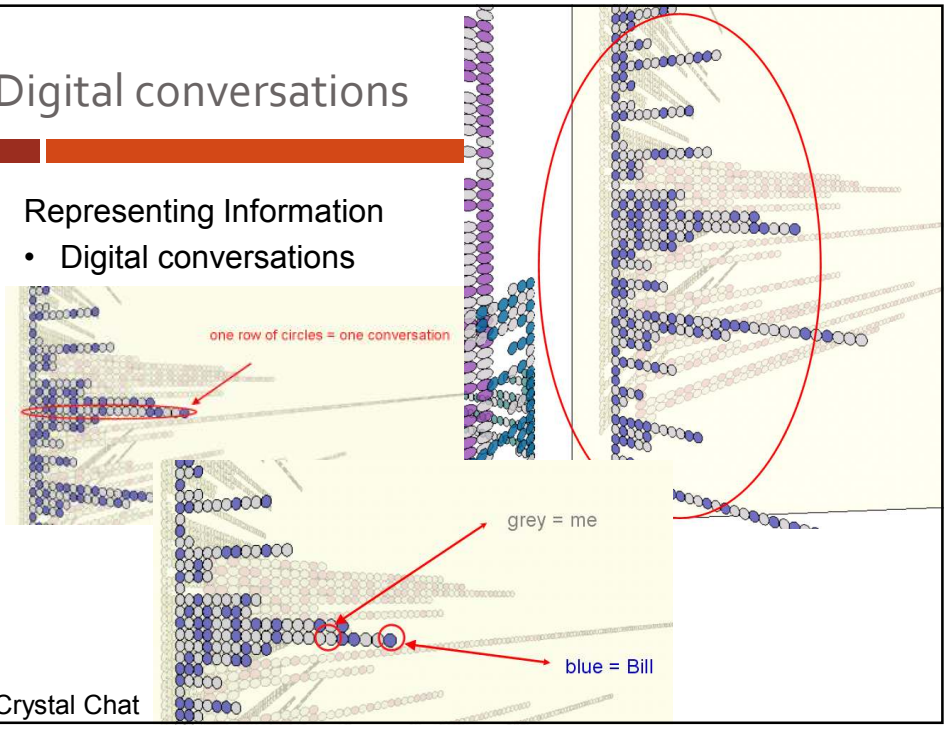
Repeated Keys

Keystrokes

Digital conversations

Representing Information

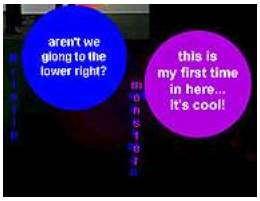
- Digital conversations



Crystal Chat

Chatting with Chat Circles

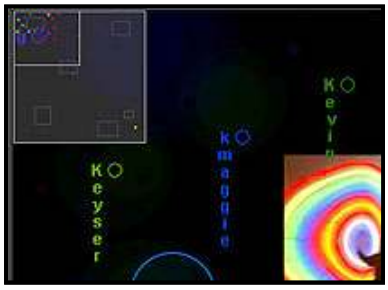
- Identity
  - Circle, Color and Name
  - Color has no meaning

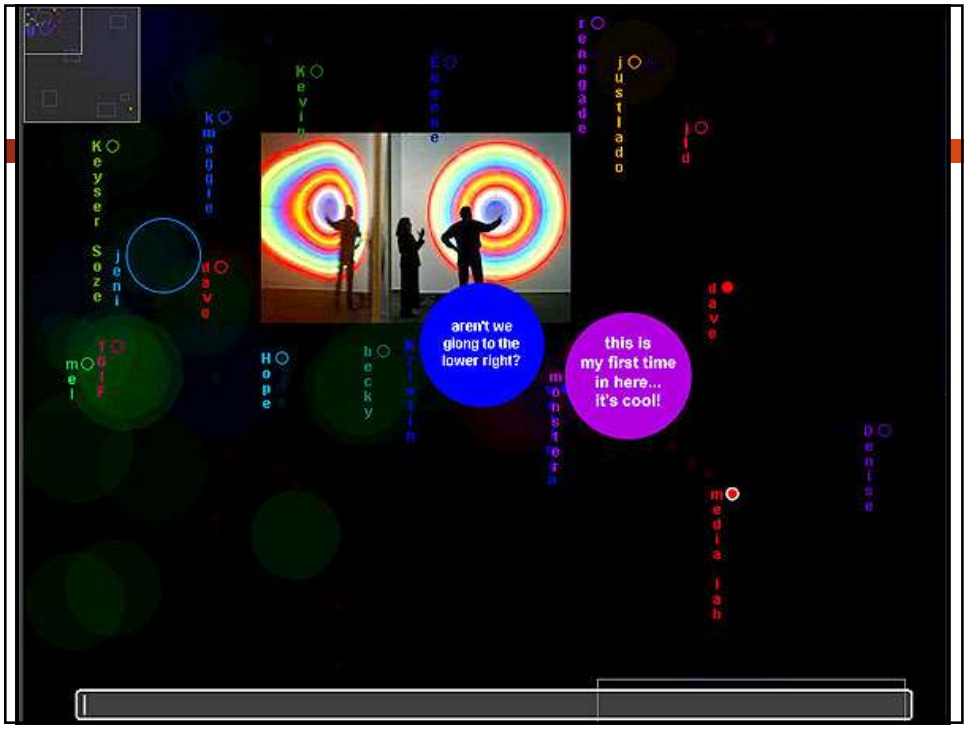


- Activity
  - Size
  - Color

Chat Circles Cont..

- Groupings
  - No rooms
  - Proximity to listen
  - Easy to spot activity

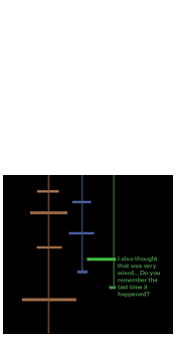


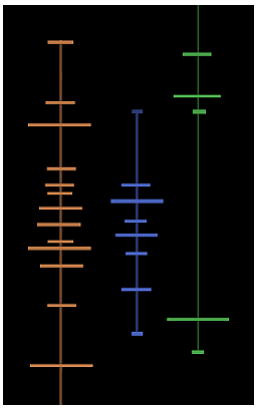


Visualizing Chat Room History

9:17 PM **me**: you're right! transparency does help
David: i don't want to talk about power points ever again
people steal ideas...
9:18 PM **me**: like who?
David: DH
9:19 PM **me**: lol
it's not stealing if you are offering it
you suggested it!

VS





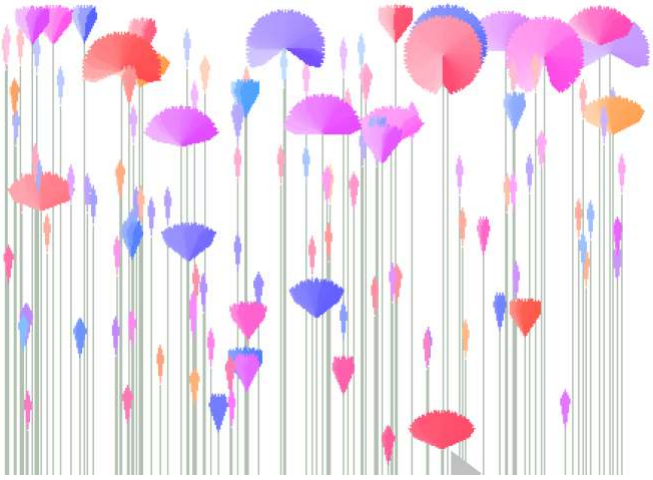
Visualizing Users

- PeopleFlower Portrait

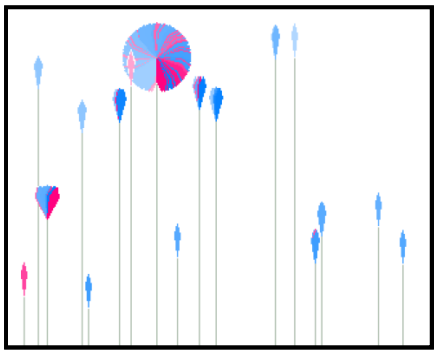
- Time of posting

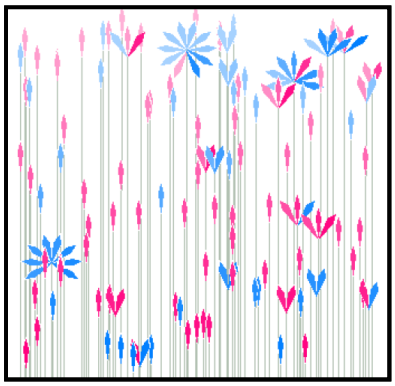
- Response

Visualizing Groups



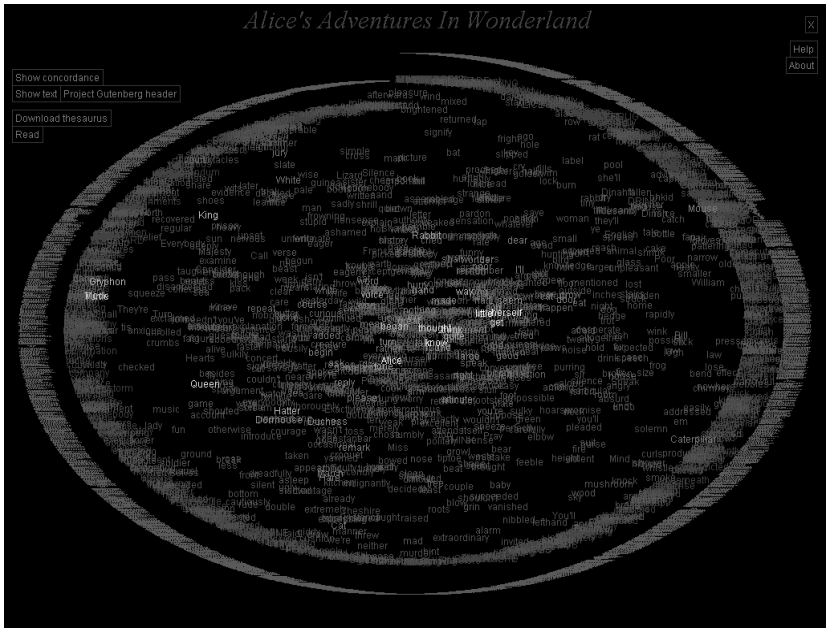
Visualizing Groups Cont…

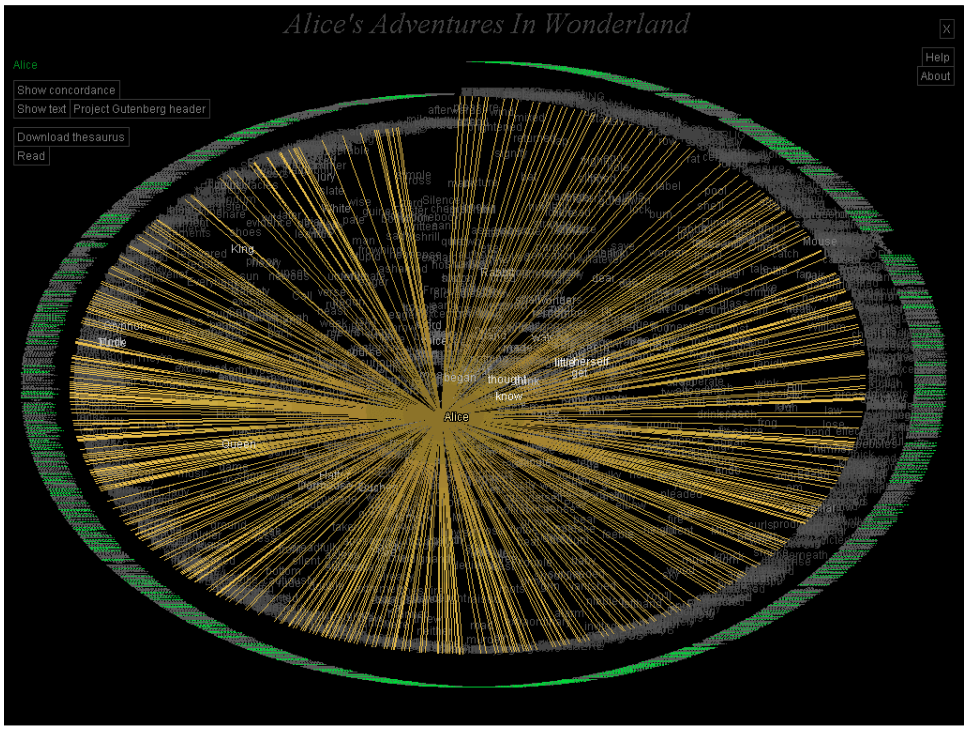


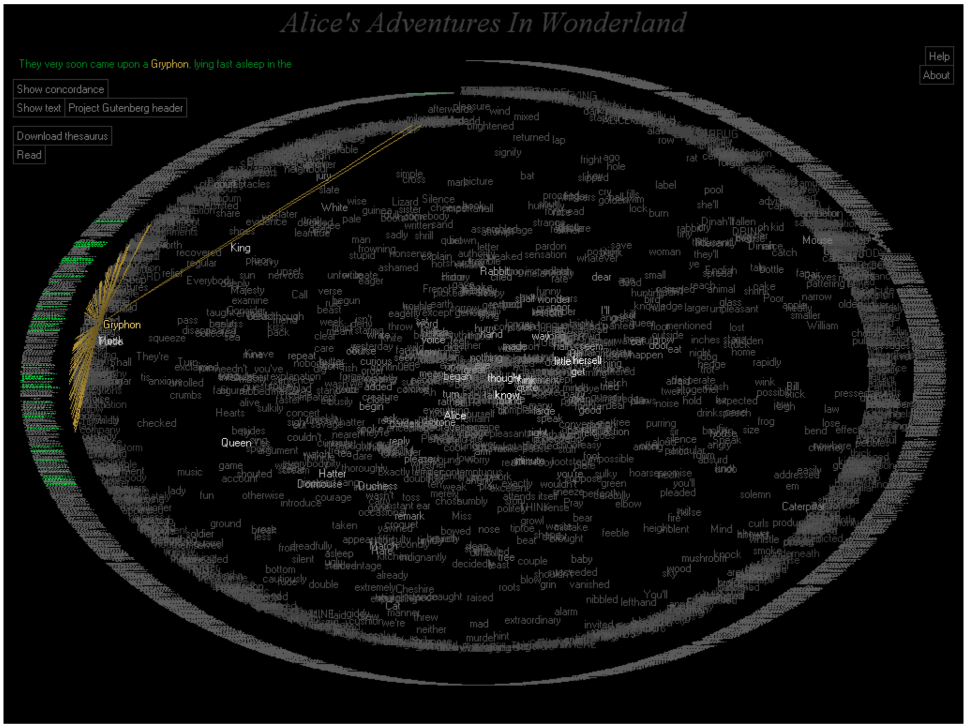


Documents

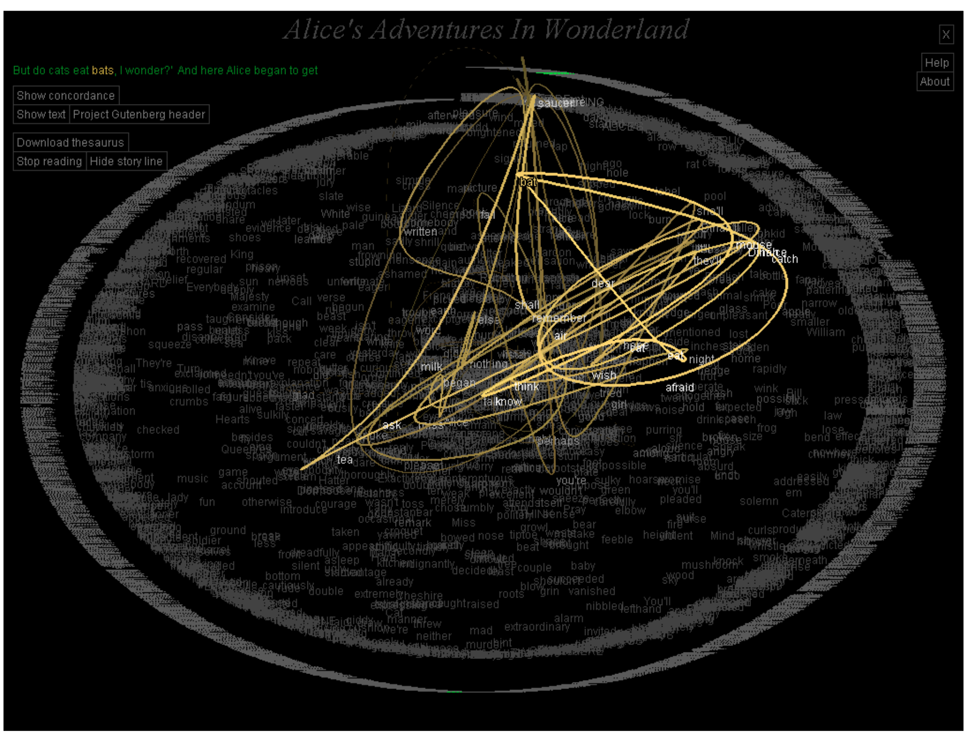
information exploration


Alice's Adventures In Wonderland
X
Help
About
Show concordance
Show text
Project Gutenberg header
Download thesaurus
Read

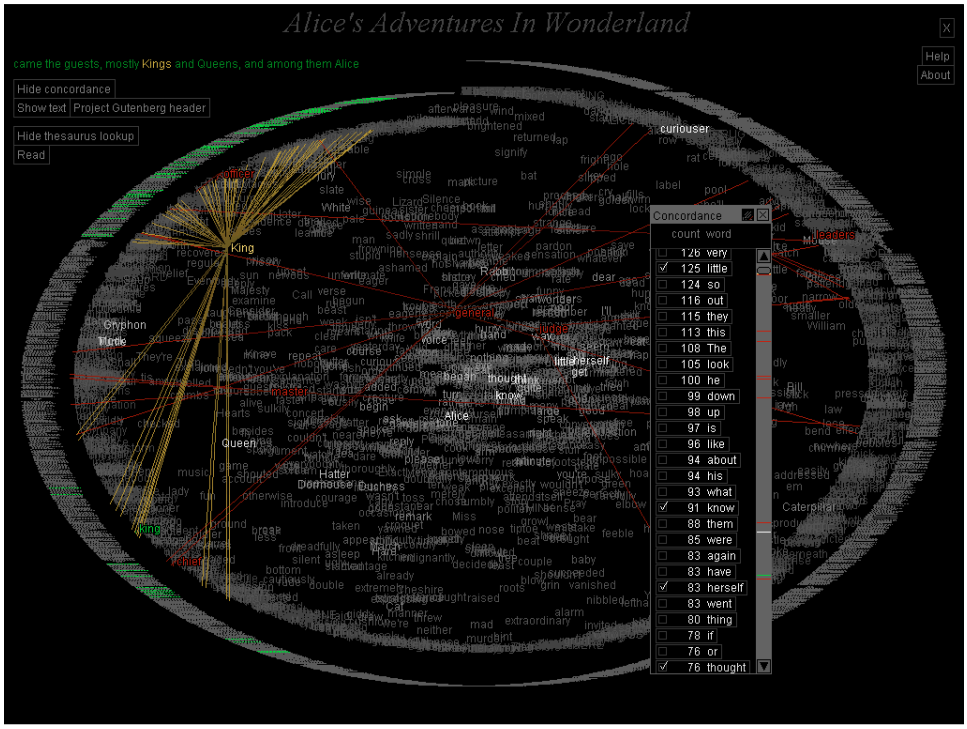

Alice's Adventures In Wonderland
Alice
X
Help
About
Show concordance
Show text
Project Gutenberg header
Download thesaurus
Read

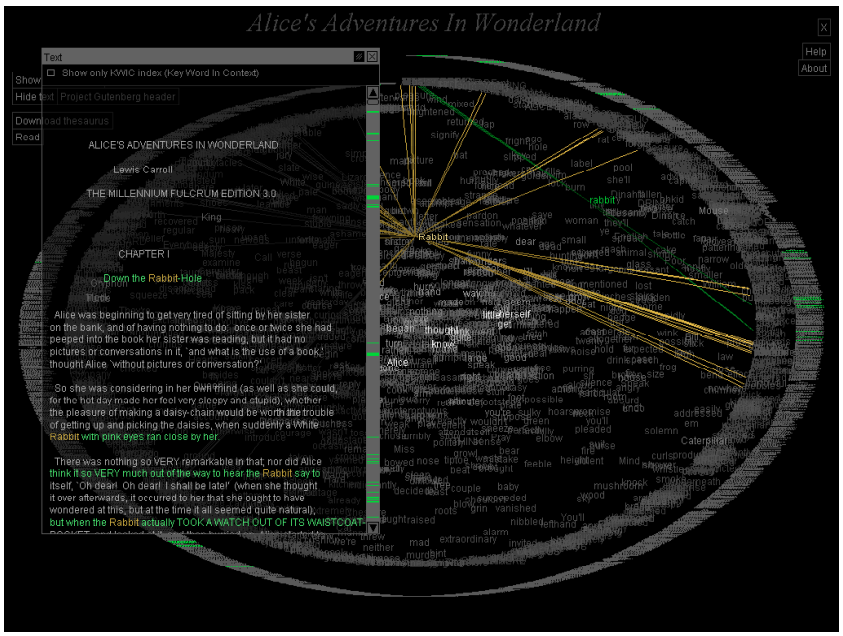


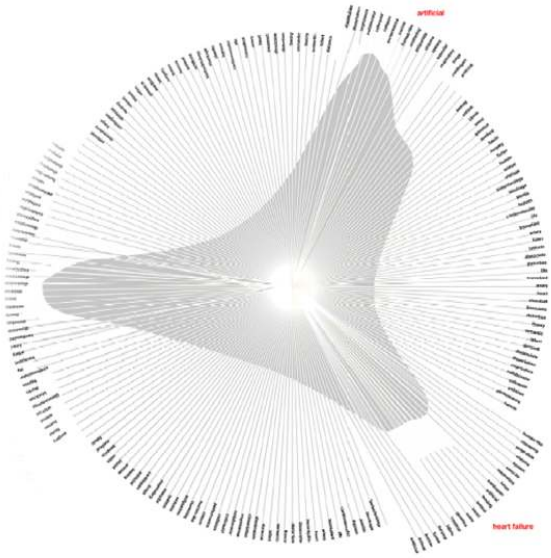
Alice's Adventures In Wonderland
They very soon came upon a Gryphon, lying fast asleep in the
Show concordance
Show text
Project Gutenberg header
Download thesaurus
Read
Help
About

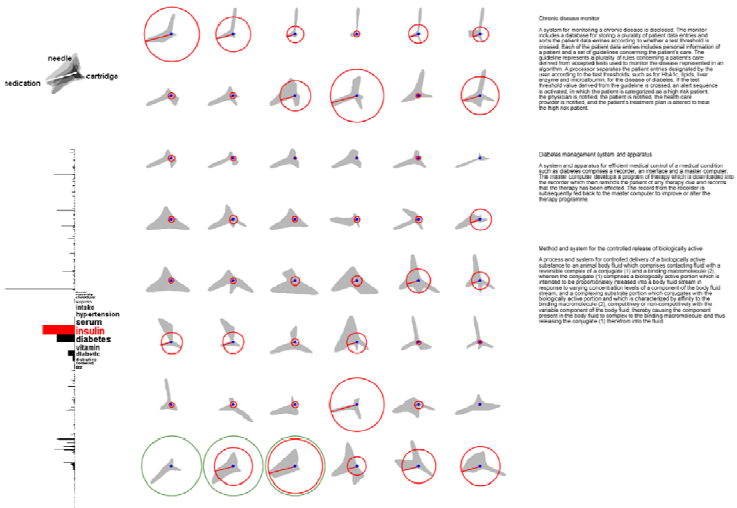


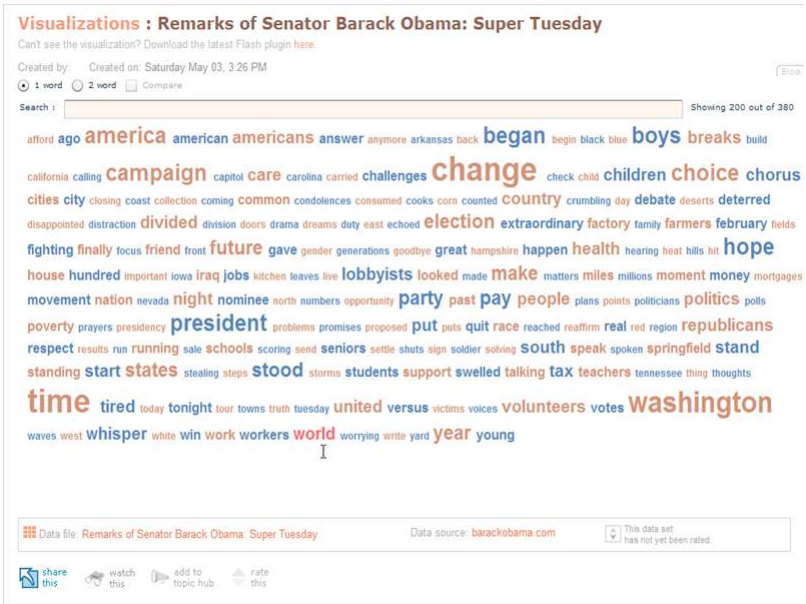
Alice's Adventures In Wonderland
But do cats eat bats, I wonder?' And here Alice began to get
Show concordance
Show text
Project Gutenberg header
Download thesaurus
Stop reading
Hide story line
Help
About


Alice's Adventures In Wonderland
came the guests, mostly Kings and Queens, and among them Alice
Hide concordance
Show text
Project Gutenberg header
Hide thesaurus lookup
Read
Help
About
Concordance
count word
126 very
125 little
124 so
116 out
115 they
113 this
108 The
105 look
100 he
99 down
98 up
97 is
96 like
94 about
94 his
93 what
91 know
88 them
85 were
83 again
83 have
83 herself
83 went
80 thing
78 if
76 or
76 thought
curiouser
King
Rabbit
Gryphon
Turtle
Alice
Queen
Hatter
Dormouse
Duchess
thought
know
little
herself


Alice's Adventures In Wonderland
Help
About
Text
Show only KWIC index (Key Word In Context)
Hide text
Project Gutenberg header
Download thesaurus
Read
ALICE'S ADVENTURES IN WONDERLAND
Lewis Carroll
THE MILLENNIUM FULCRUM EDITION 3.0
CHAPTER I
Down the Rabbit-Hole
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'
So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.
There was nothing so VERY remarkable in that; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually TOOK A WATCH OUT OF ITS WAISTCOAT-
rabbit
Rabbit
thought
know
little
herself
Alice

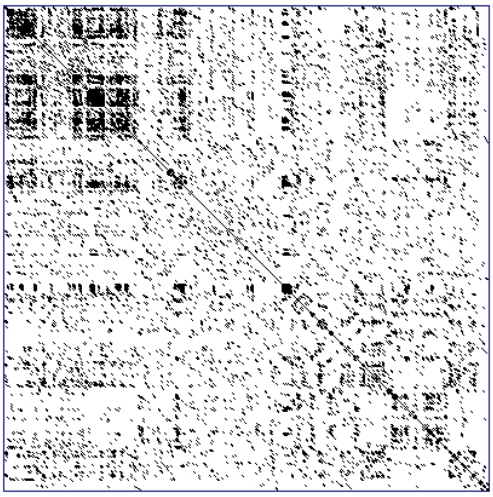

artificial
heart failure


needle
cartridge
medication
intake
hyp-ertension
serum
insulin
diabetes
vitamin
diabetic

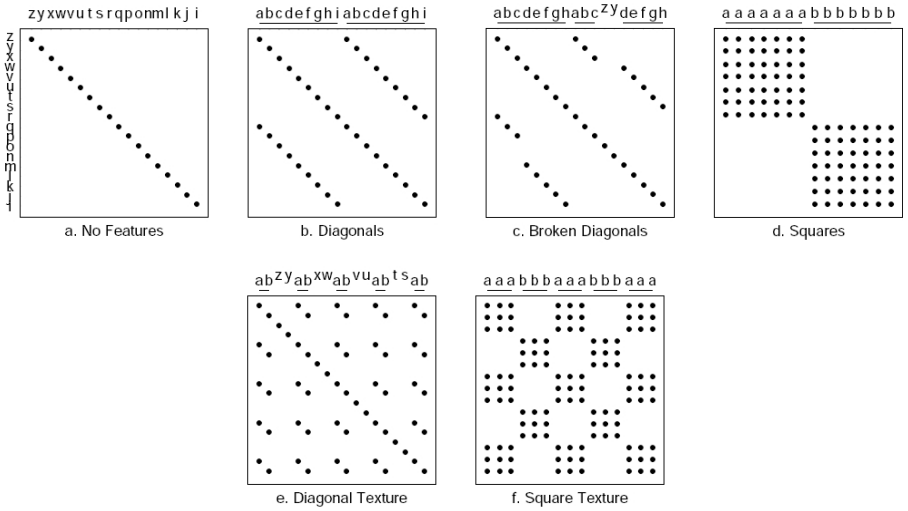


Visualizing structure in strings

Dot Plot



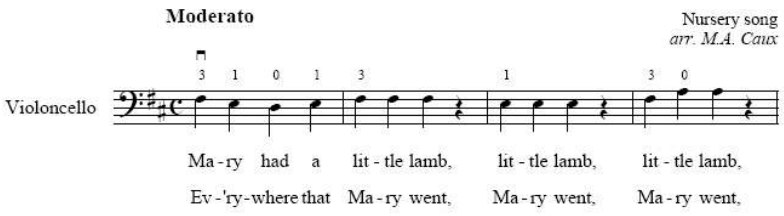
Dot Plot Patterns



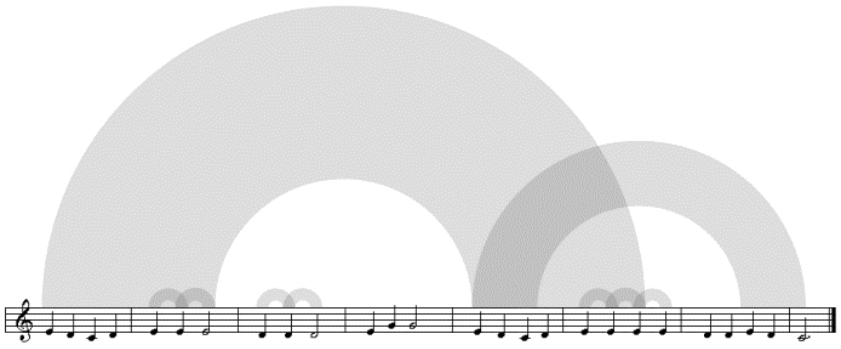
Basic Definition

1. Identical
2. Non-overlapping
3. Consecutive
4. Maximal

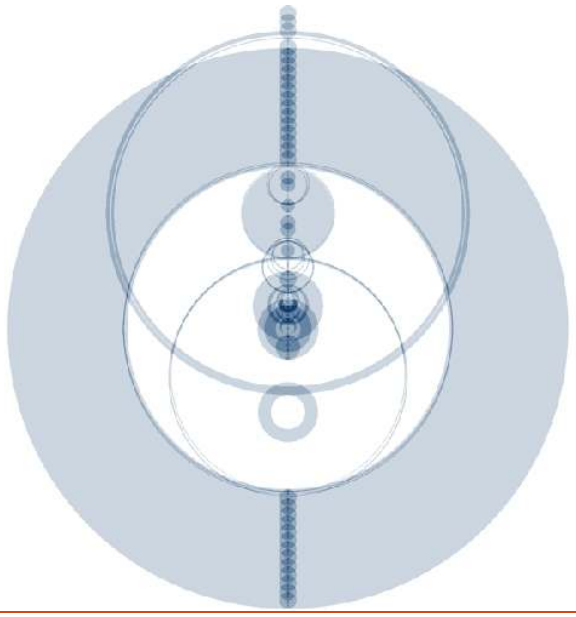
Mary Had a Little Lamb



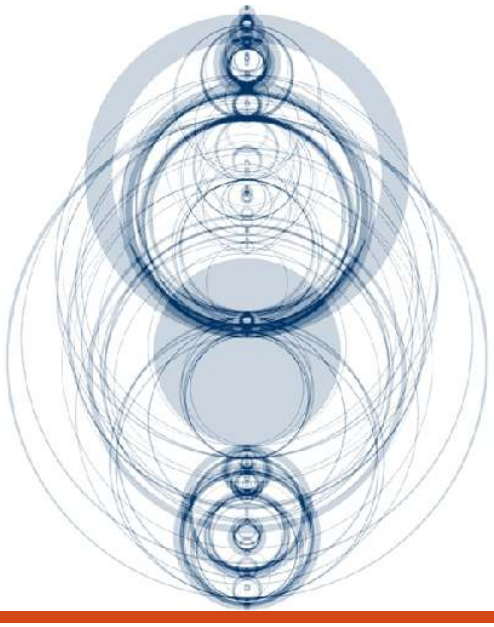
Mary Had a Little Lamb



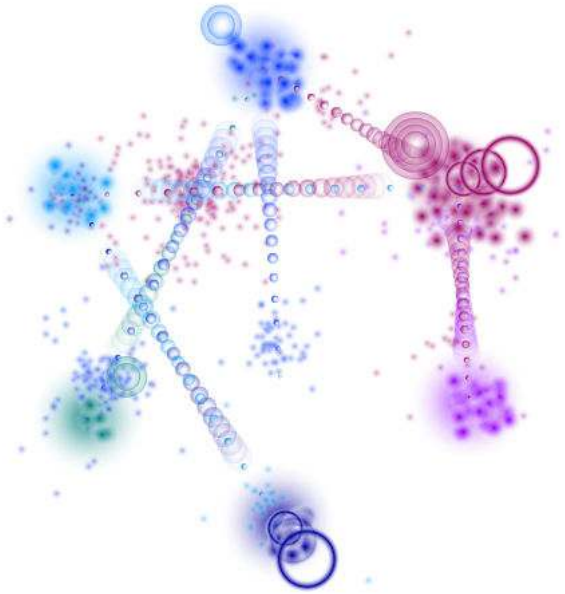
Philip Glass - Koyaanisqatsi



Vivaldi - Autumn

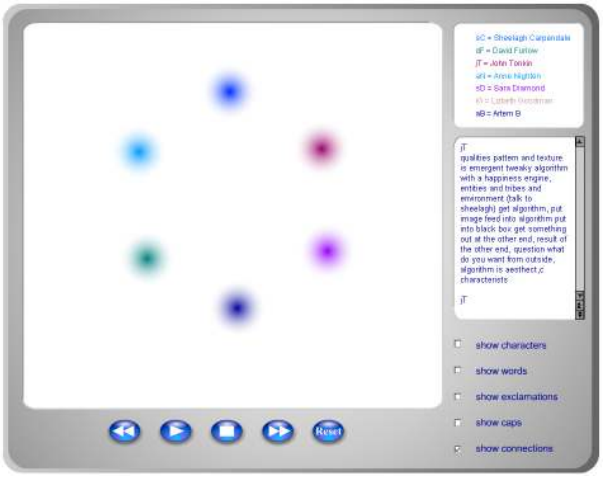


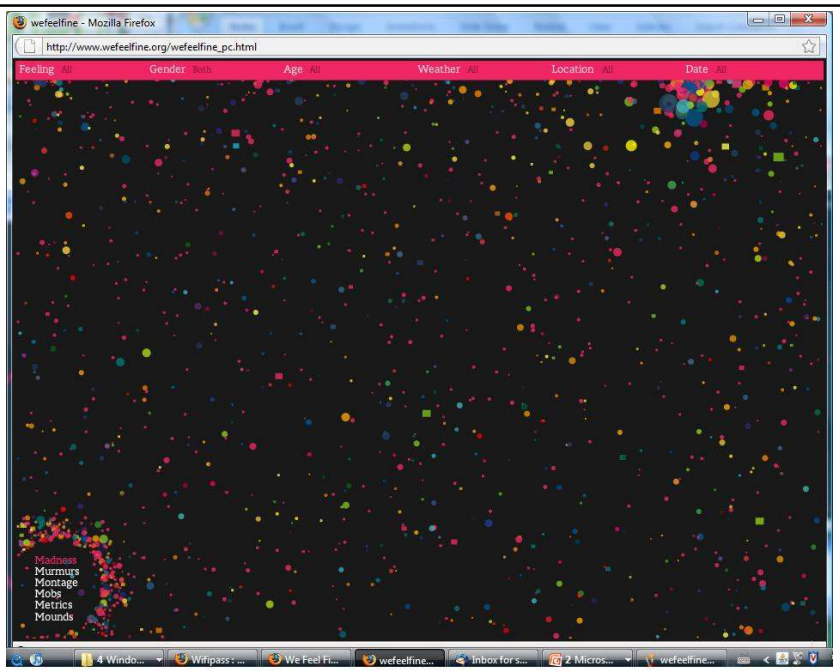
Streaming Data

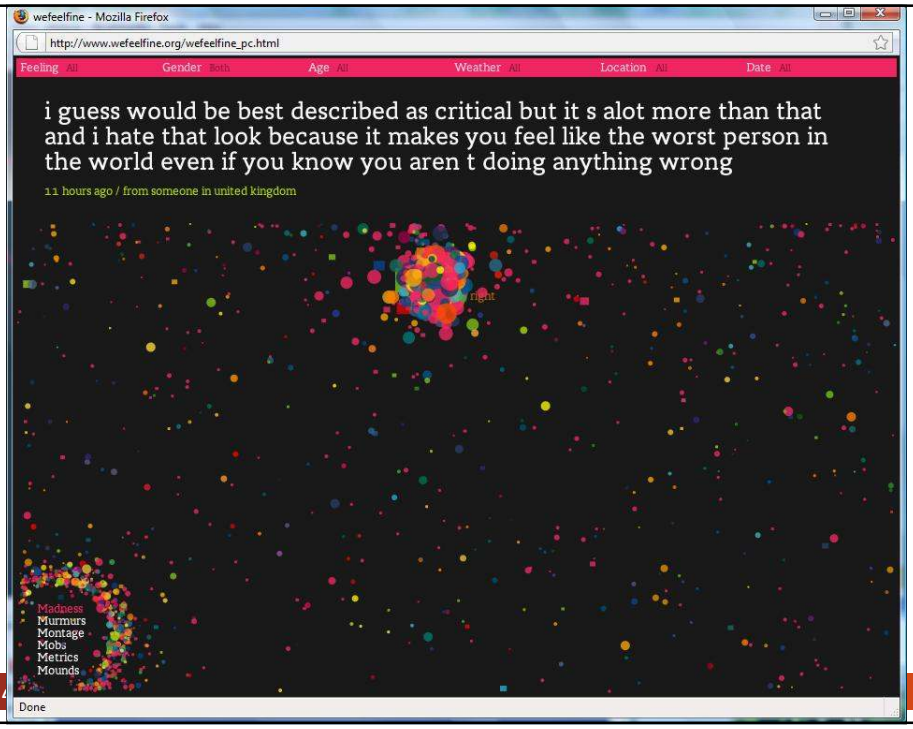


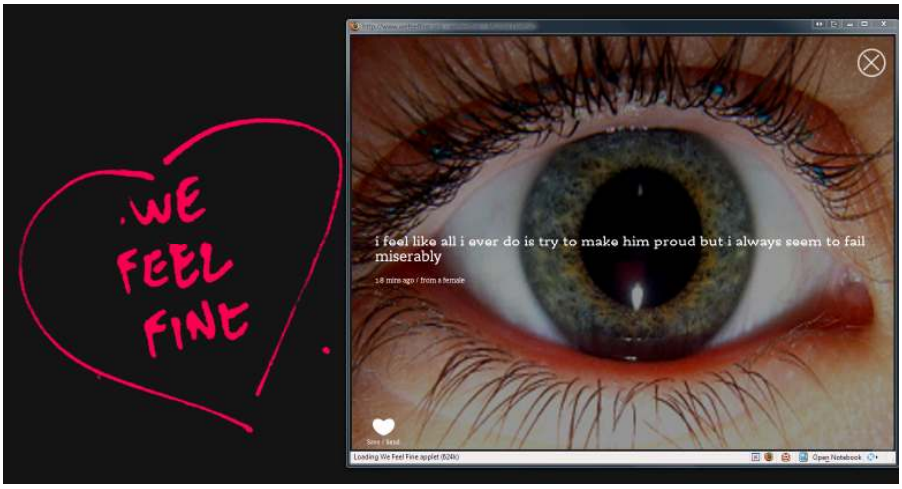
Annie Tat

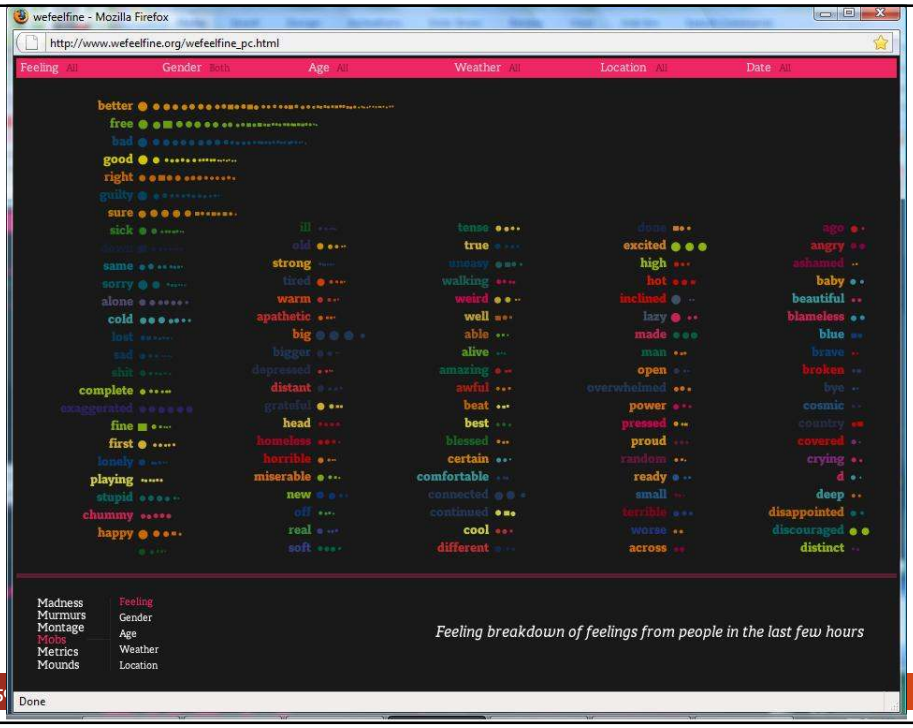
Demonstration

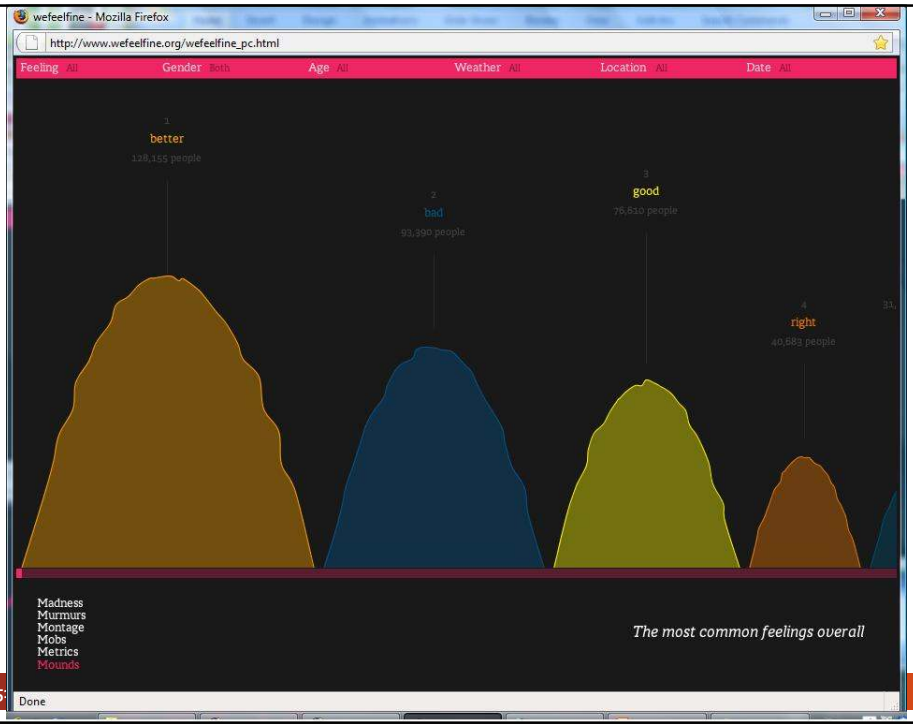


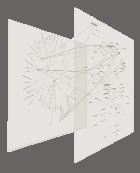

wefeelfine - Mozilla Firefox
http://www.wefeelfine.org/wefeelfine_pc.html
Feeling All
Gender Both
Age All
Weather All
Location All
Date All
Madness
Murmurs
Montage
Mobs
Metrics
Mounds


wefeelfine - Mozilla Firefox
http://www.wefeelfine.org/wefeelfine_pc.html
Feeling All
Gender Both
Age All
Weather All
Location All
Date All
i guess would be best described as critical but it s alot more than that and i hate that look because it makes you feel like the worst person in the world even if you know you aren t doing anything wrong
11 hours ago / from someone in united kingdom
Madness
Murmurs
Montage
Mobs
Metrics
Mounds
Done


.WE
FEEL
FINE
i feel like all i ever do is try to make him proud but i always seem to fail miserably


wefeelfine - Mozilla Firefox
http://www.wefeelfine.org/wefeelfine_pc.html
Feeling All
Gender Both
Age All
Weather All
Location All
Date All
Madness
Murmurs
Montage
Mobs
Metrics
Mounds
Feeling
Gender
Age
Weather
Location
Feeling breakdown of feelings from people in the last few hours
Done


wefeelfine - Mozilla Firefox
http://www.wefeelfine.org/wefeelfine_pc.html
Feeling All
Gender Both
Age All
Weather All
Location All
Date All
1
better
128,155 people
2
bad
93,390 people
3
good
76,810 people
4
right
40,683 people
Madness
Murmurs
Montage
Mobs
Metrics
Mounds
The most common feelings overall
Done




Interactive Visualization for Computational Linguistics

ESSLLI 2009

2 Visualization Design

Graphical integrity

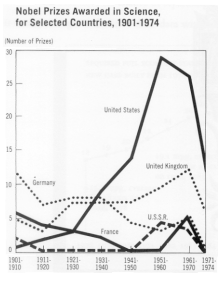
- Graphics can be a powerful communication tool
- Lies and falsehoods are possible
- Much focus on this 'how to lie with maps' or 'statistics'

Extrapolation

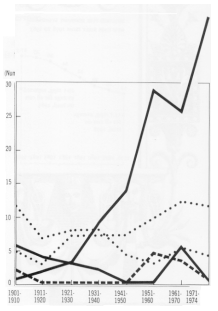
1st seven intervals are 10 years

The last interval is 4 years

Gives a false sense of decline



National Science Foundation, Science Indicators, 1974 (Washington D.C., 1976), p.15, (Tufte, 1983, p60)



Accurate data for the next 10 years

Visual Area and Numerical measure

Use of area or volume to portray 1D data can be confusing

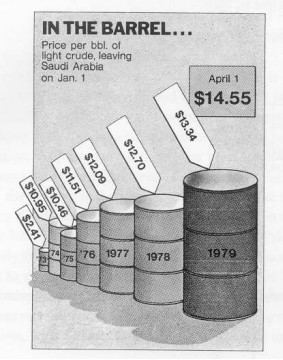
There are considerable ambiguities of how people perceive area (2D) and then convert that to 1D data

Even more so with volumes (3D)

By surface area lie factor is 9.4

By volume – lie factor is 59.4

The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.



Time, Apr. 9, 1979 p.1, (Tufte, 1983, p62)

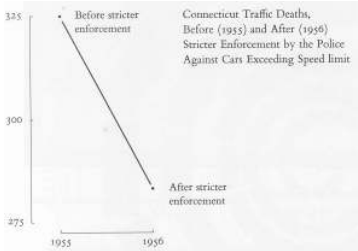
Context is Essential

Graphics must not quote data out of context

Data sparse graphics should provoke suspicion

Graphics often lie by omission

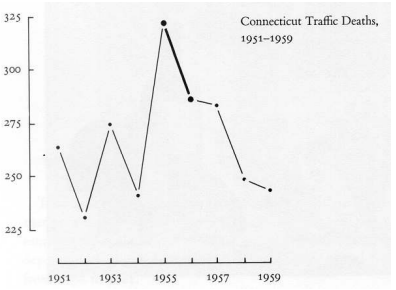
Nearly all important questions are left unanswered by this graphic



Context is Essential

Graphics must quote data out of context

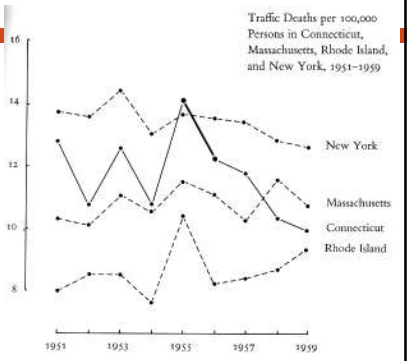
A few more data points tell a more complete story



Context is Essential

Graphics must not quote data out of context

Comparisons with adjacent states give more context



Donald T. Campbell and H. Laurence Ross, "The Connecticut Crackdown on Speeding: Time Series Data in Quasi-Experimental Analysis," in Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems* (Reading, Mass., 1970), 115–125.

Graphical Integrity - Summary

- 'The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the graphic itself. Label important events.
- Show data variation, not design variation.
- In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
- Graphics must not quote data out of context.'

(Tufte, 1983, p77)

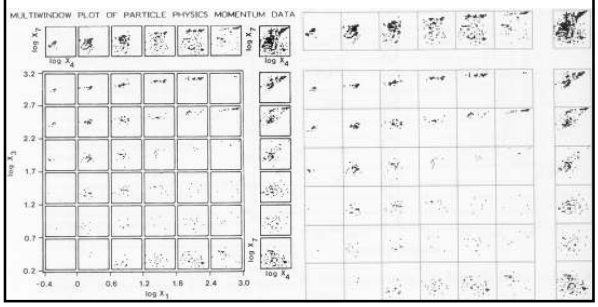
Tufte's Graphical Theory

- 1st graphical excellence (1st set)
- 2nd graphical disasters (2nd set)

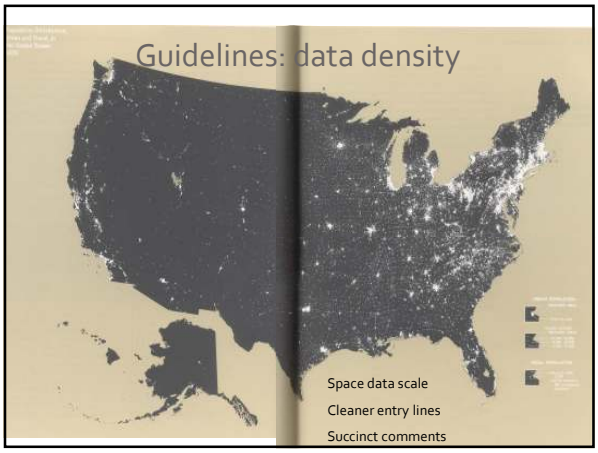
- guidelines (this slide set)
  1. Minimize ink while maximizing data ink
  2. Data density
  3. Avoid chartjunk
  4. Create multifunctioning graphics
  5. Make use of parallelism (small multiples)
  6. Consider 3D carefully
  7. Consider colour carefully

Guidelines: Minimize ink while maximizing data ink

heavy grid lines –non data ink – actually cause irritation – chartjunk



Guidelines: data density

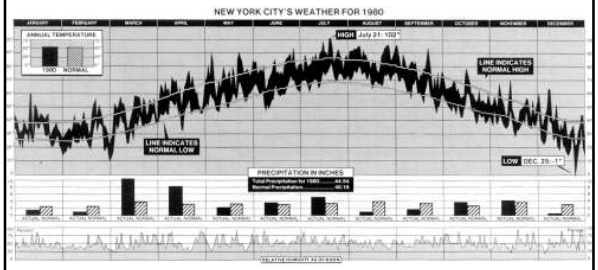


Space data scale

Cleaner entry lines

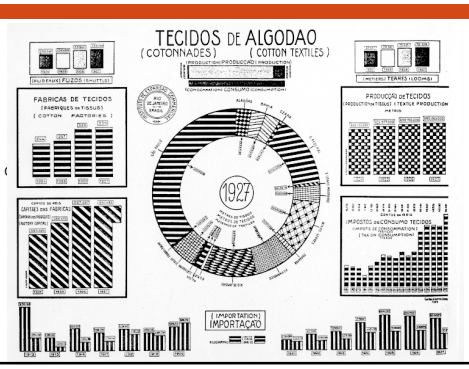
Succinct comments

Guidelines: data density



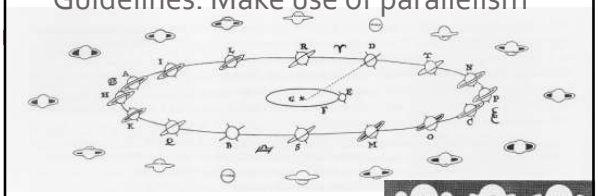
New York Times, (Jan. 1981), p.32, (Tufte, 1883, p.30)

Guidelines: Avoid chartjunk

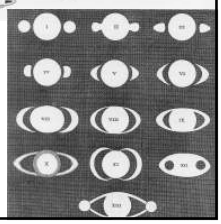


Guidelines: Create multifunctioning graphics

Guidelines: Make use of parallelism

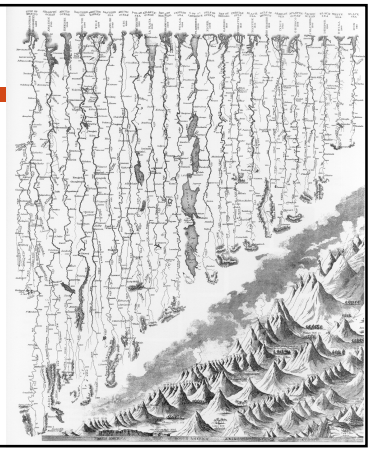


- Small Multiple from Huygens' *Systema Saturnium* 1659.
- Inner ellipse is earth's orbit, outer Saturn's orbit, outer most floating images depict Saturn as viewed from earth, explaining previous idea about Saturn's shape

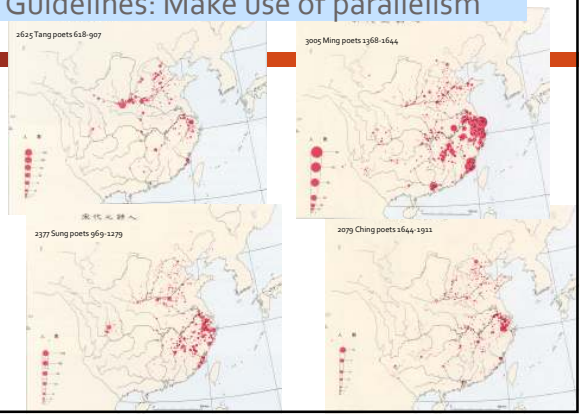


Guidelines: Make use of parallelism

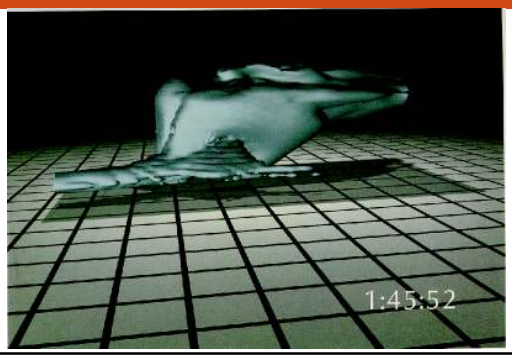
- J. H. Colton
- Johnson's New Illustrated Family Atlas with Physical Geography, 1864



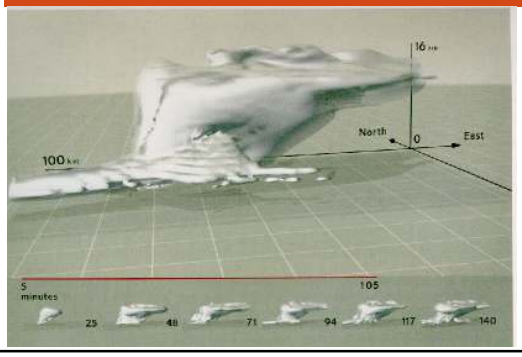
Guidelines: Make use of parallelism



Small Multiples: Showing Time and Change



Small Multiples: Showing Time and Change

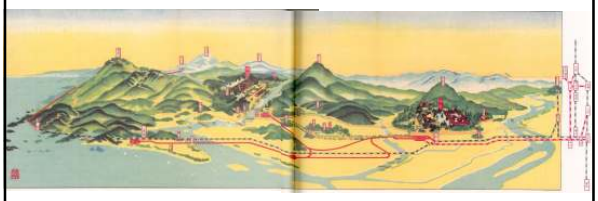


Guidelines: Consider 3D carefully

- Why 3D?
  - The world is 3D – 2D graphics flatten data

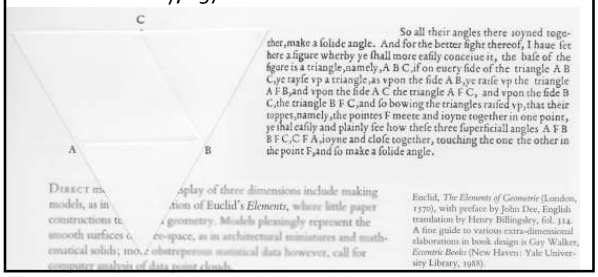
Guidelines: Consider 3D carefully

- 3D graphic excellence



Guidelines: Consider 3D carefully

- 3D graphic excellence
- Actual 3D with paper – Euclid, The Elements of Geometry, 1570



So all their angles there ioyned together, make a ſolide angle. And for the better light thereof, I haue ſet here a figure wherby ye ſhall more eaſily conceiue it, the baſe of the figure is a triangle, namely, A B C, if on euery ſide of the triangle A B C, ye rayſe vp a triangle, as vpon the ſide A B, ye raiſe vp the triangle A F B, and vpon the ſide A C the triangle A F C, and vpon the ſide B C, the triangle B F C, and ſo bowing the triangles raiſed vp, that their toppes, namely, the pointes F meete and ioyne together in one point, ye ſhal eaſily and plainly ſee how theſe three ſuperficiall angles A F B B F C, C F A, ioyne and cloſe together, touching the one the other in the point F, and ſo make a ſolide angle.

C

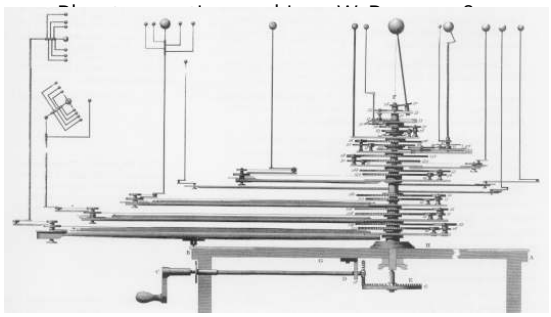
A

B

Euclid, *The Elements of Geometrie* (London, 1570), with preface by John Dee, English translation by Henry Billingsley, fol. 314. A fine guide to various extra-dimensional elaborations in book design is Guy Walker, *Eccentric Books* (New Haven: Yale University Library, 1988).

Guidelines: Consider 3D carefully

- 3D graphic excellence

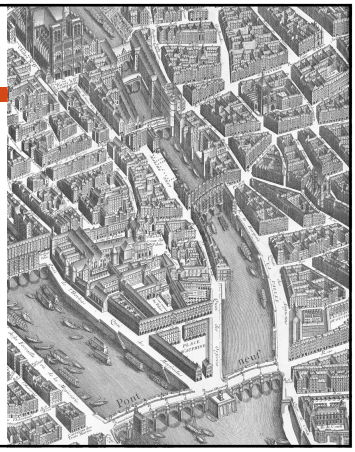


Guidelines: Consider 3D carefully

- 2 paintings on silk depicting Deshima Island, a view from the Bay (top), a view from Nagasaki (bottom), circa 1860

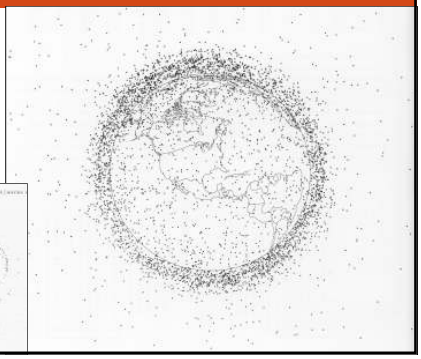
Guidelines: Consider 3D carefully

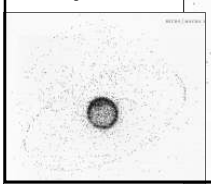
- 3D graphic excellence
- M. E. Turgot, L. Bretez, Plan de Paris (Paris 1739)



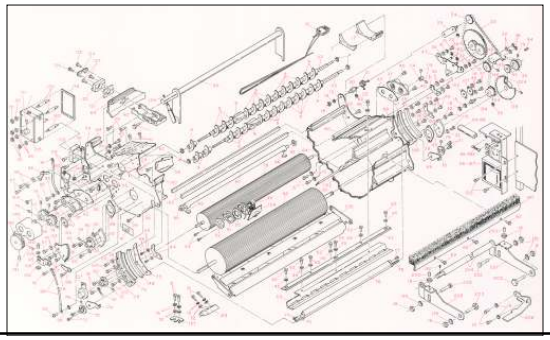
Guidelines: Consider 3D carefully

- 3D graphic excellence
- Some 7000 pieces of space debris – points not to scale but each at least as big as 8x10 inches – 1987
- Micro and macro readings





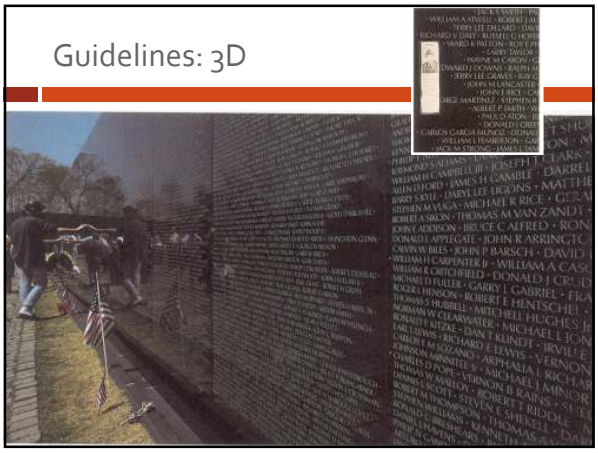
Guidelines: Consider 3D carefully



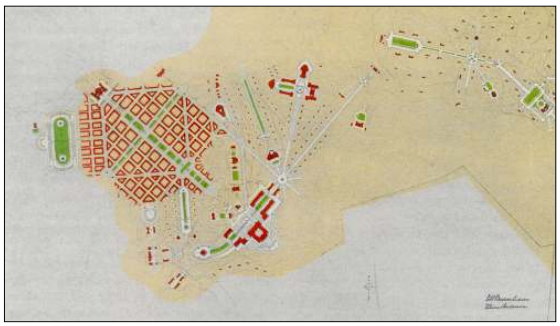
Guidelines: 3D

3D or 2D?

Guidelines: 3D

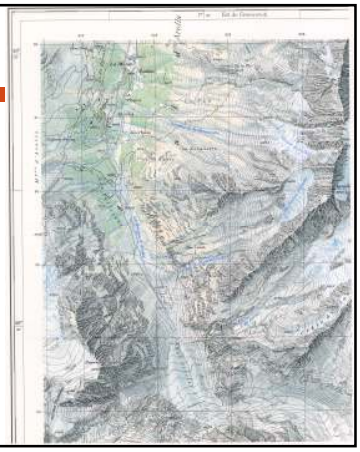


Guidelines: Consider colour carefully

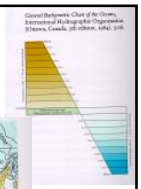


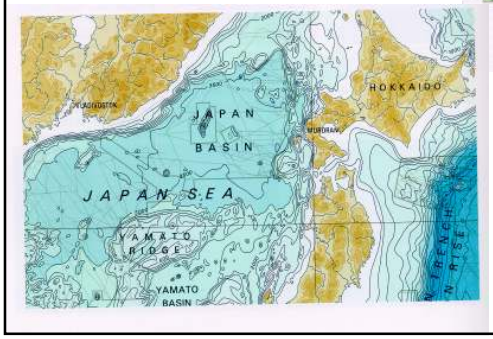
Guidelines: Consider colour carefully

- To avoid over powering and garish colours
- Use least possible distinguishable difference



Guidelines: Consider colour carefully

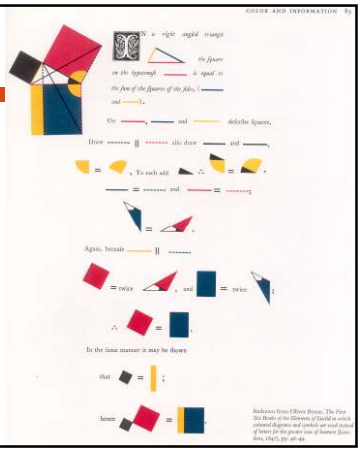




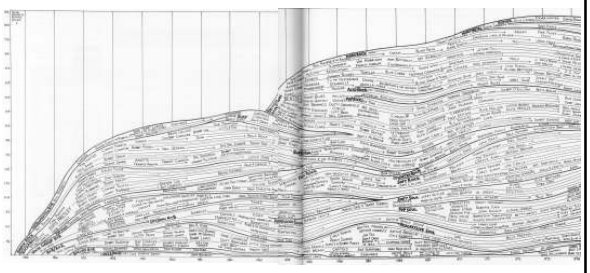
Guidelines:

Consider colour carefully

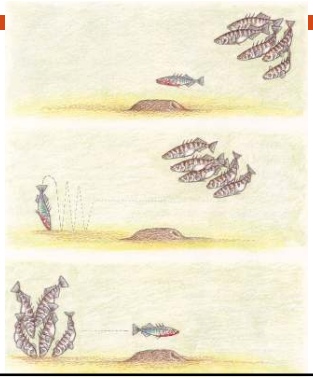
- Geometry with coloured diagrammatic explanations



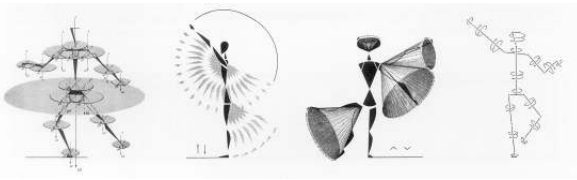
Guidelines: narratives of space and time



Guidelines: narratives of space and time

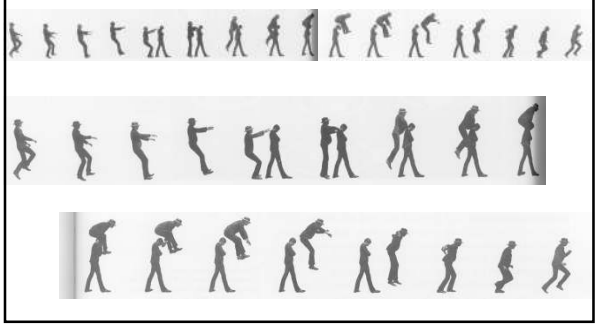


Guidelines: narratives of space and time

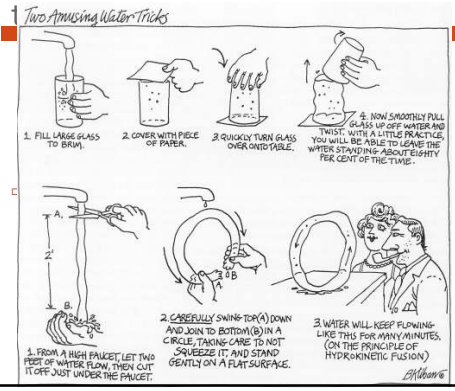


Guidelines: narratives of space and time

Guidelines: narratives of space and time



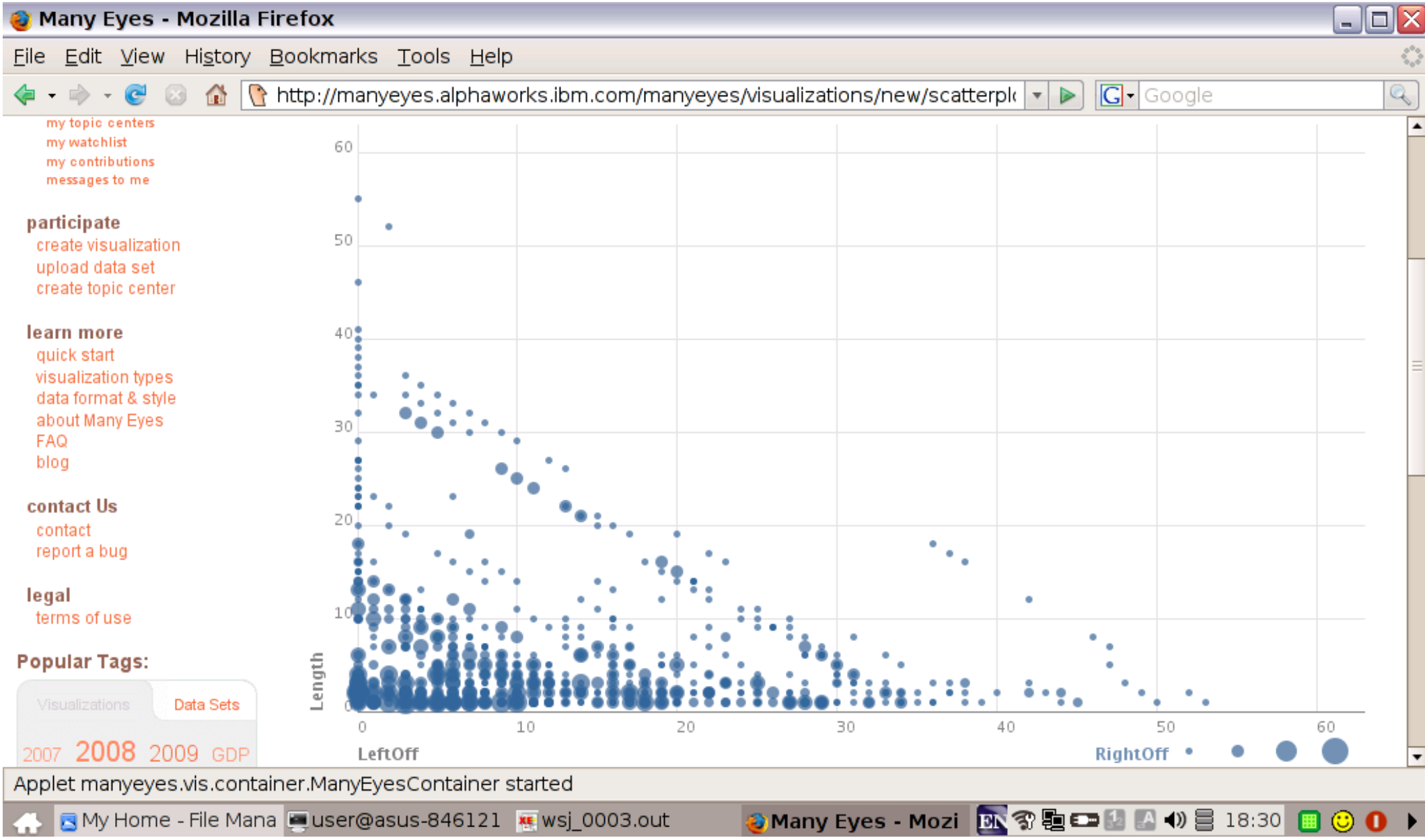
Guidelines: narratives of space and



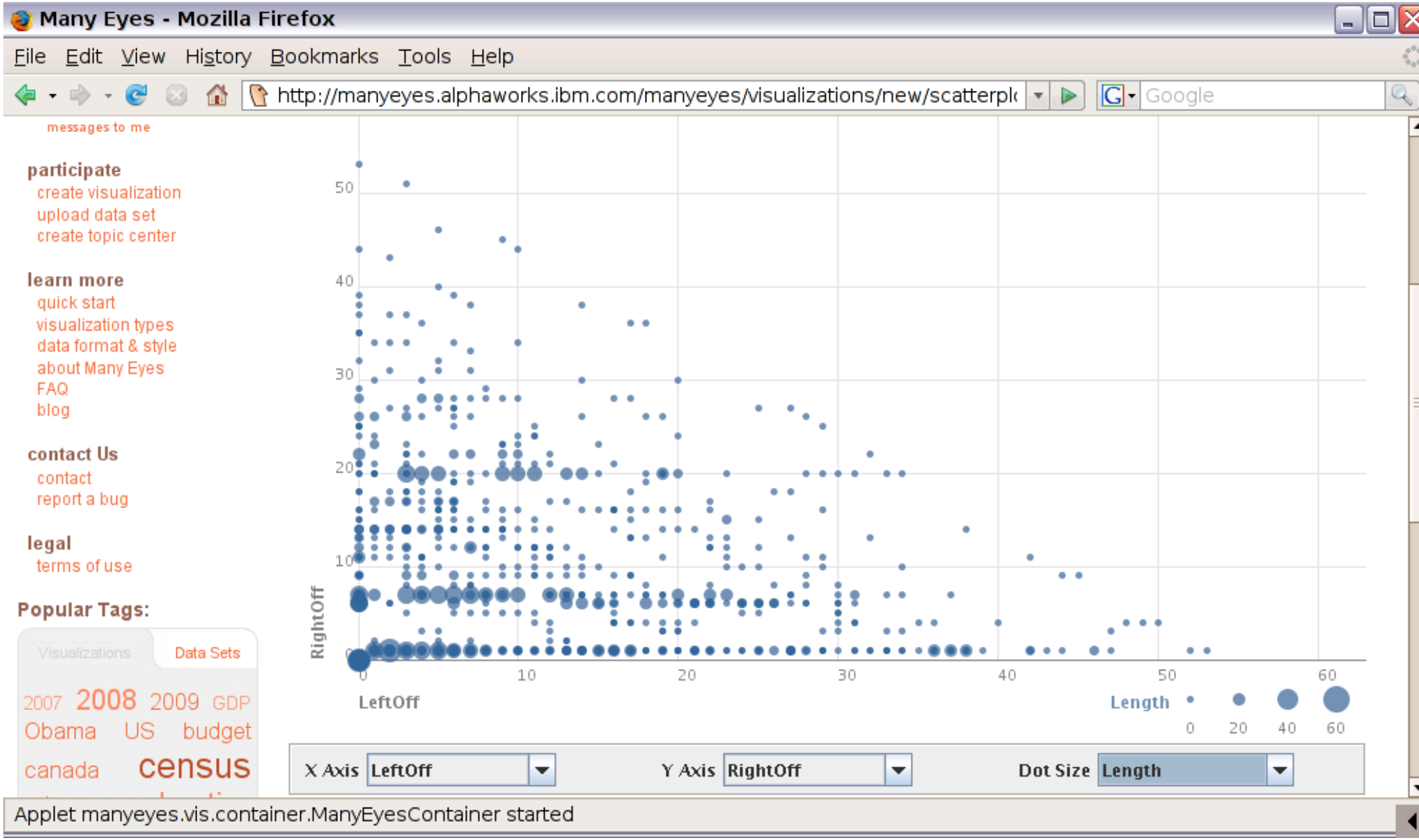
Graphical Integrity - Summary

- **Representation of numbers in the graphic should be directly proportional to data numbers**
- **Labeling**
- **Show data variation, not design variation.**
- **Standardized units of monetary measurement are nearly always better than nominal units.**
- **Information-carrying dimensions should not exceed the number of dimensions in the data.**
- **Graphics must not quote data out of context**

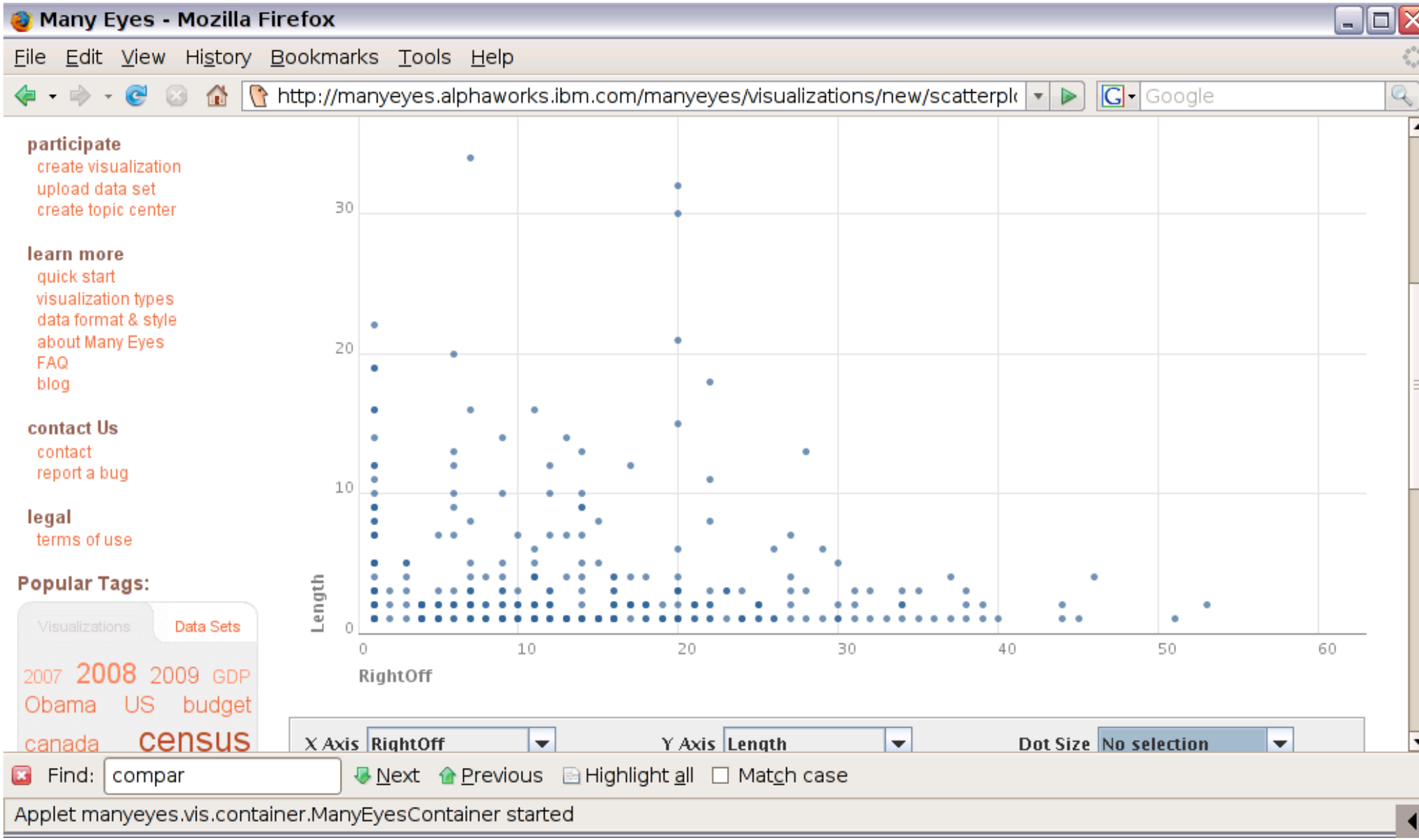
(Tufte, 1983, p77)



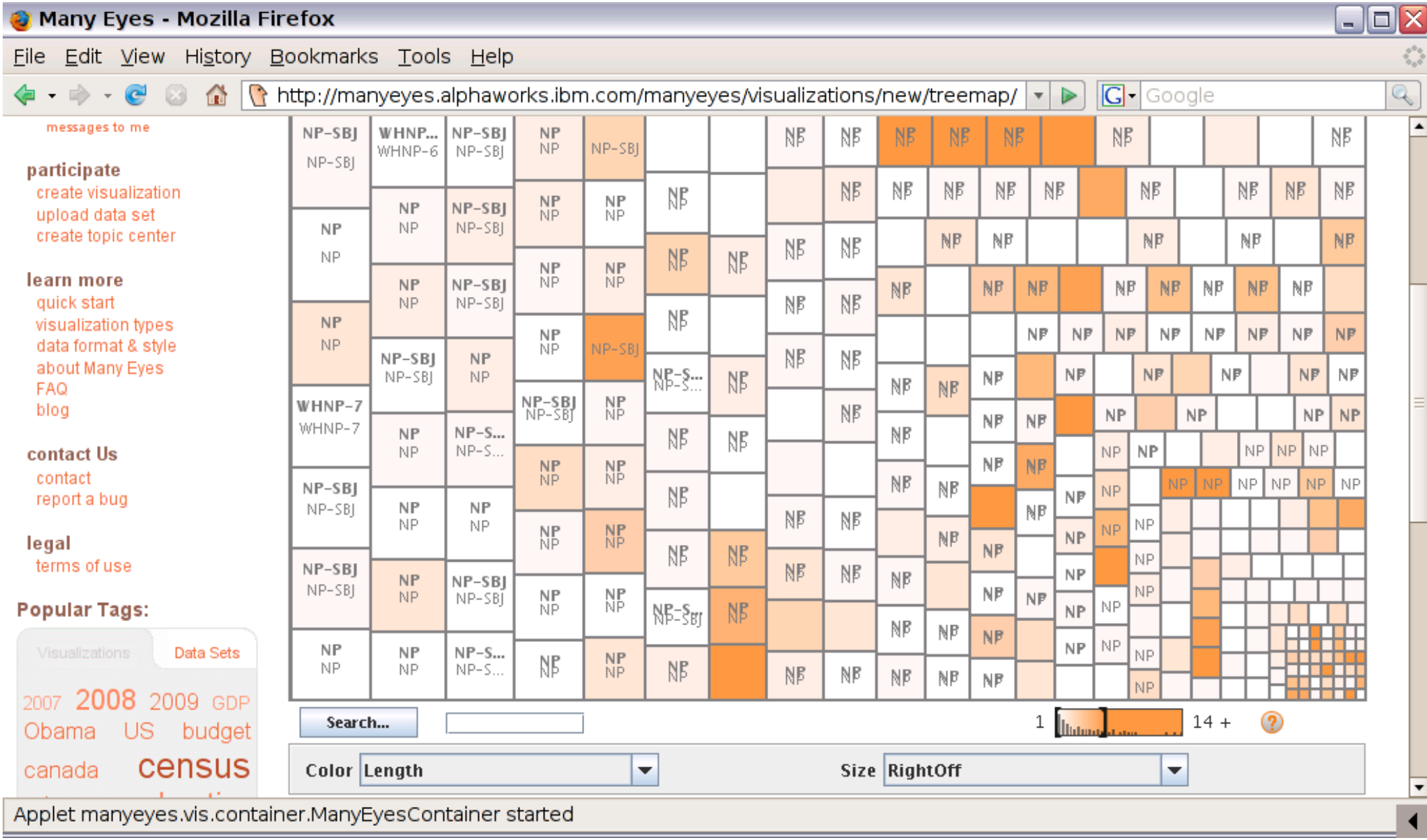
Scatterplot: left offset x length x right offset



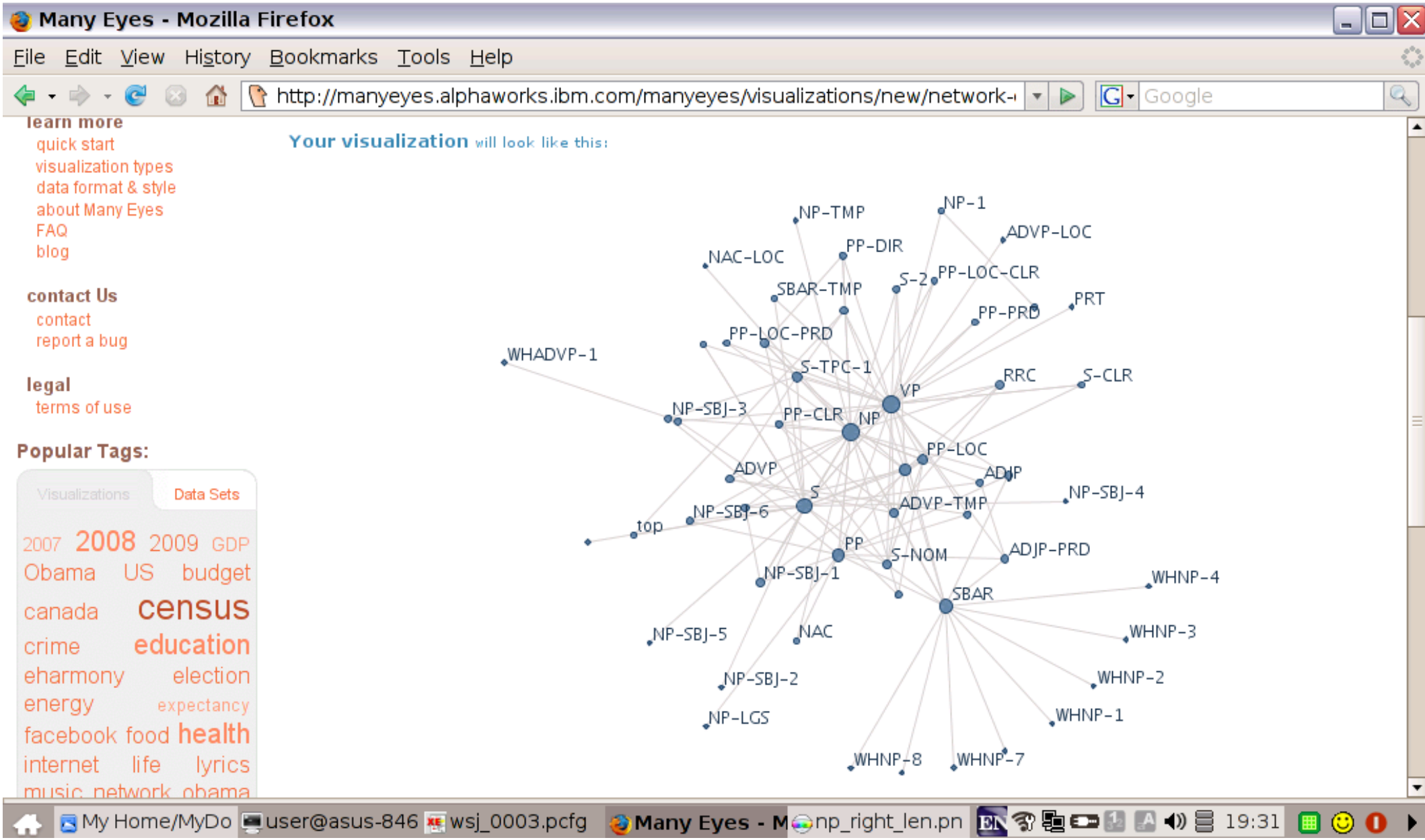
Scatterplot: left offset x right offset x length



Scatterplot: NPs by right offset x length

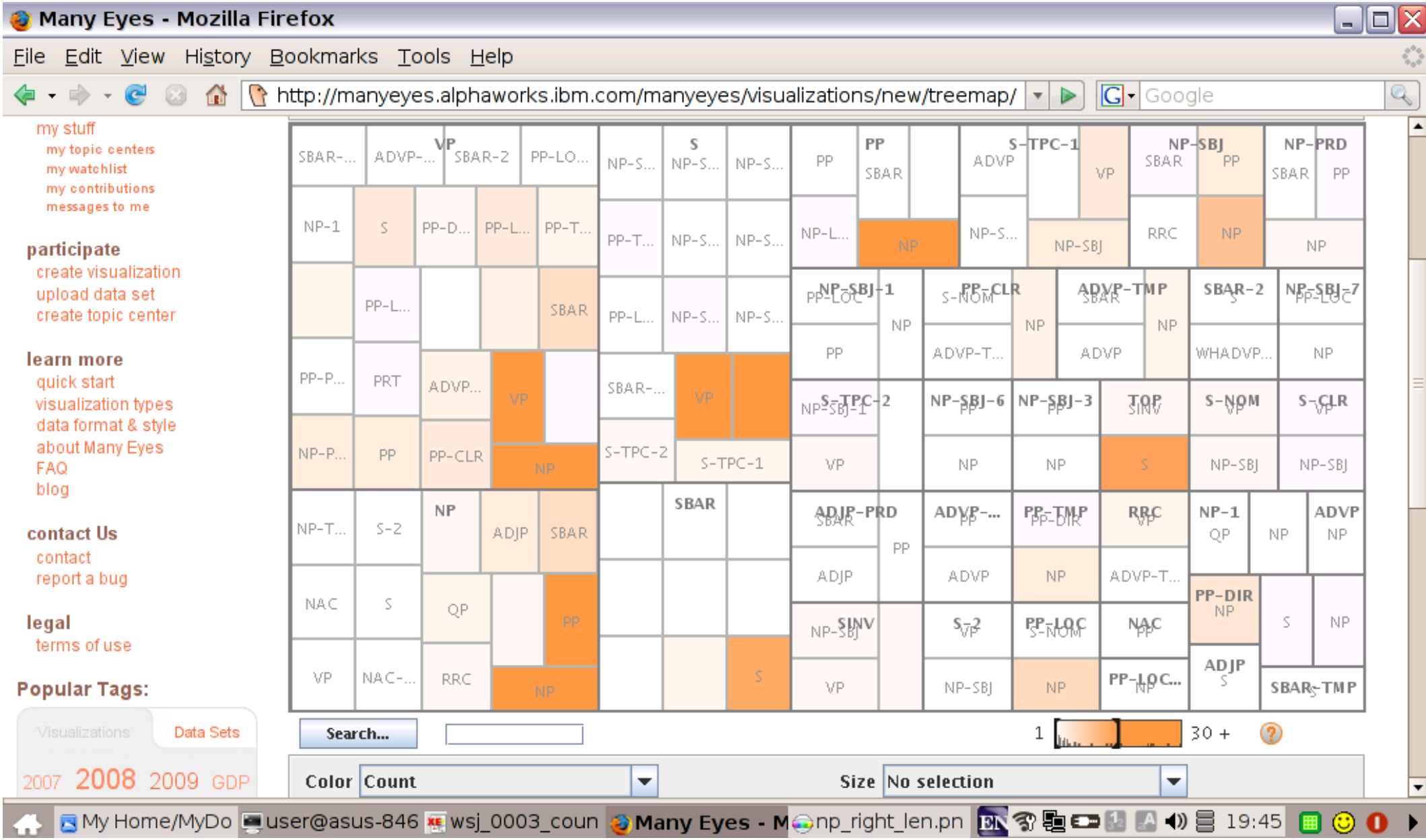


Treemap: NPs by length x right offset

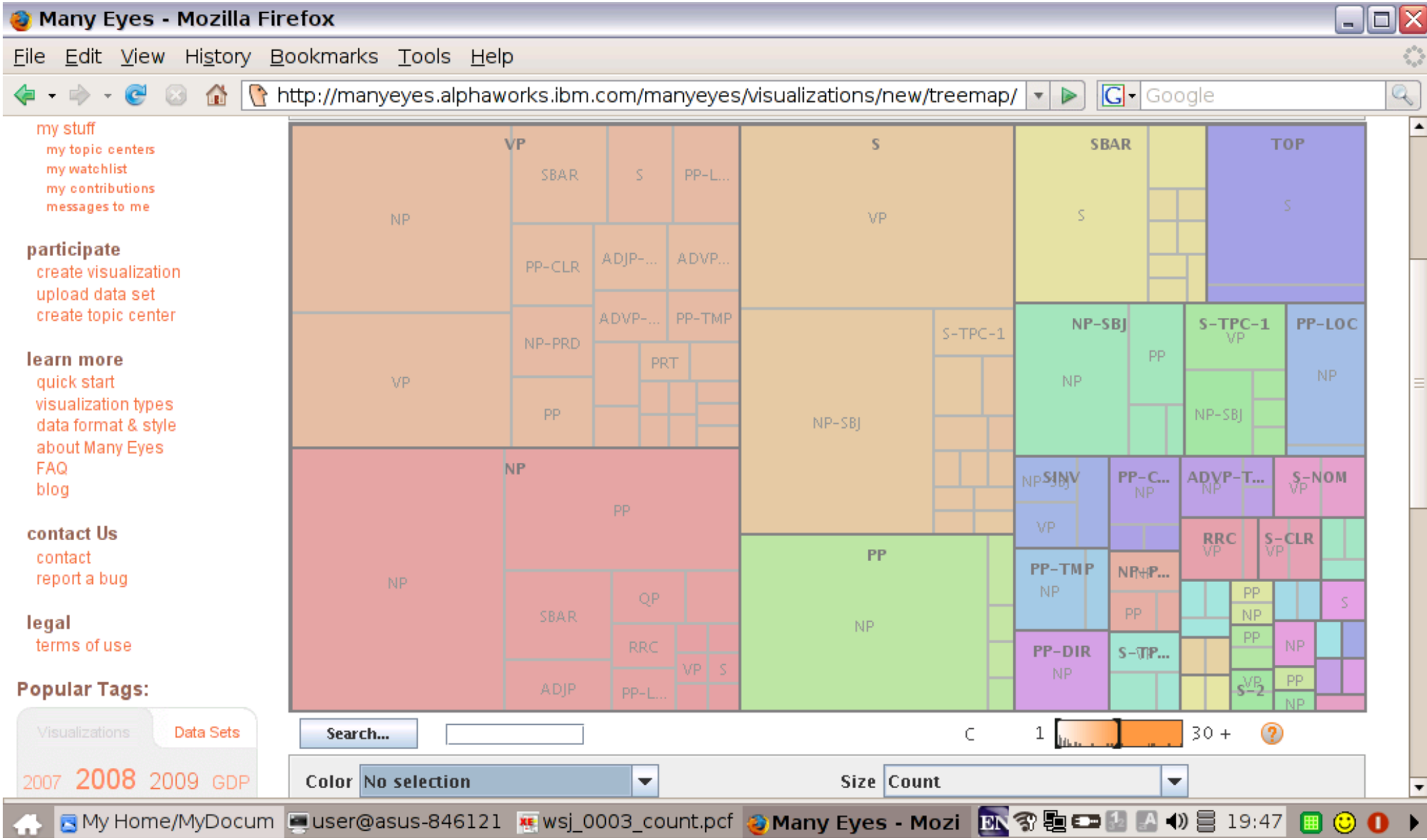


Network graph: parent-child dependencies

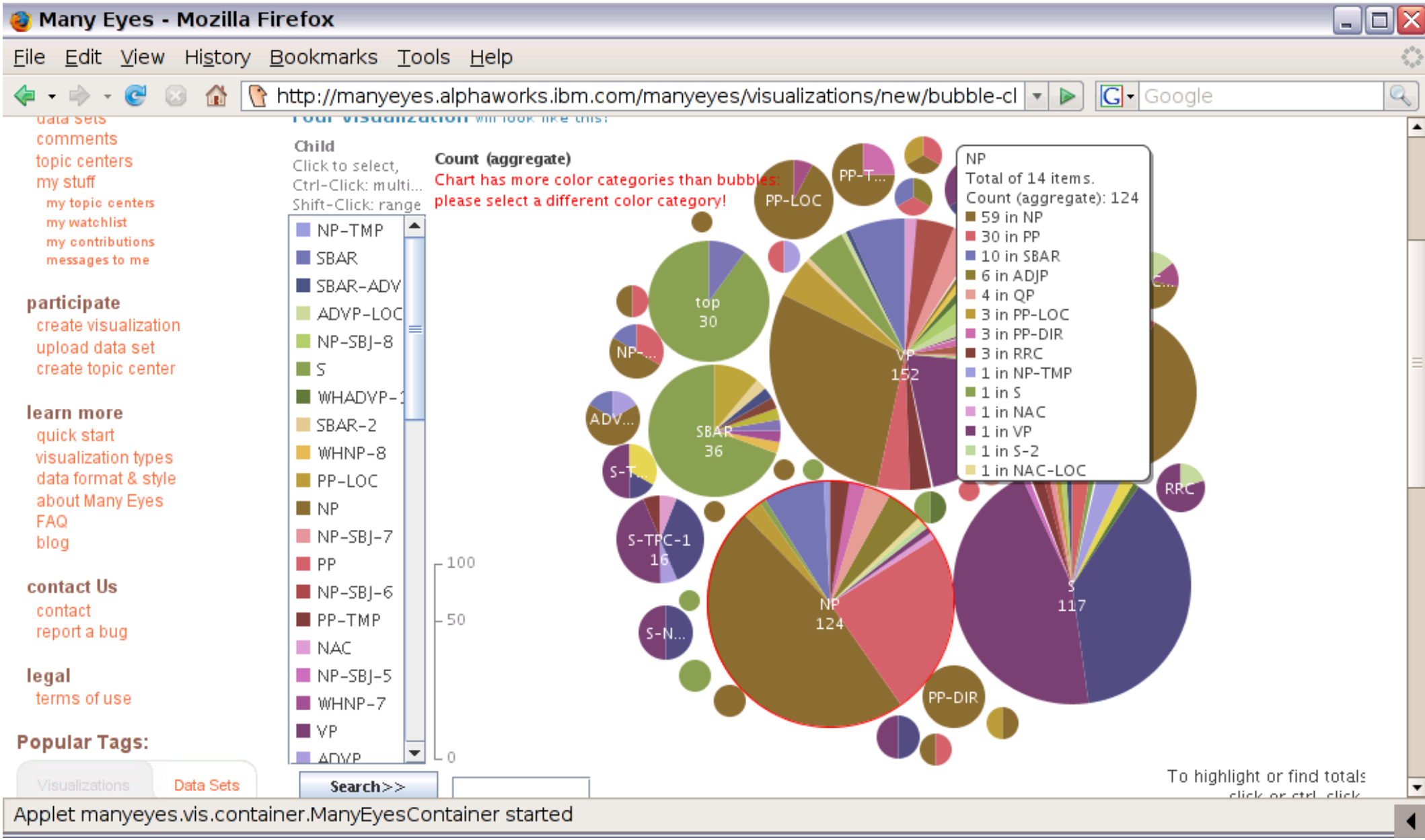
Matrix chart: parent-child dependencies



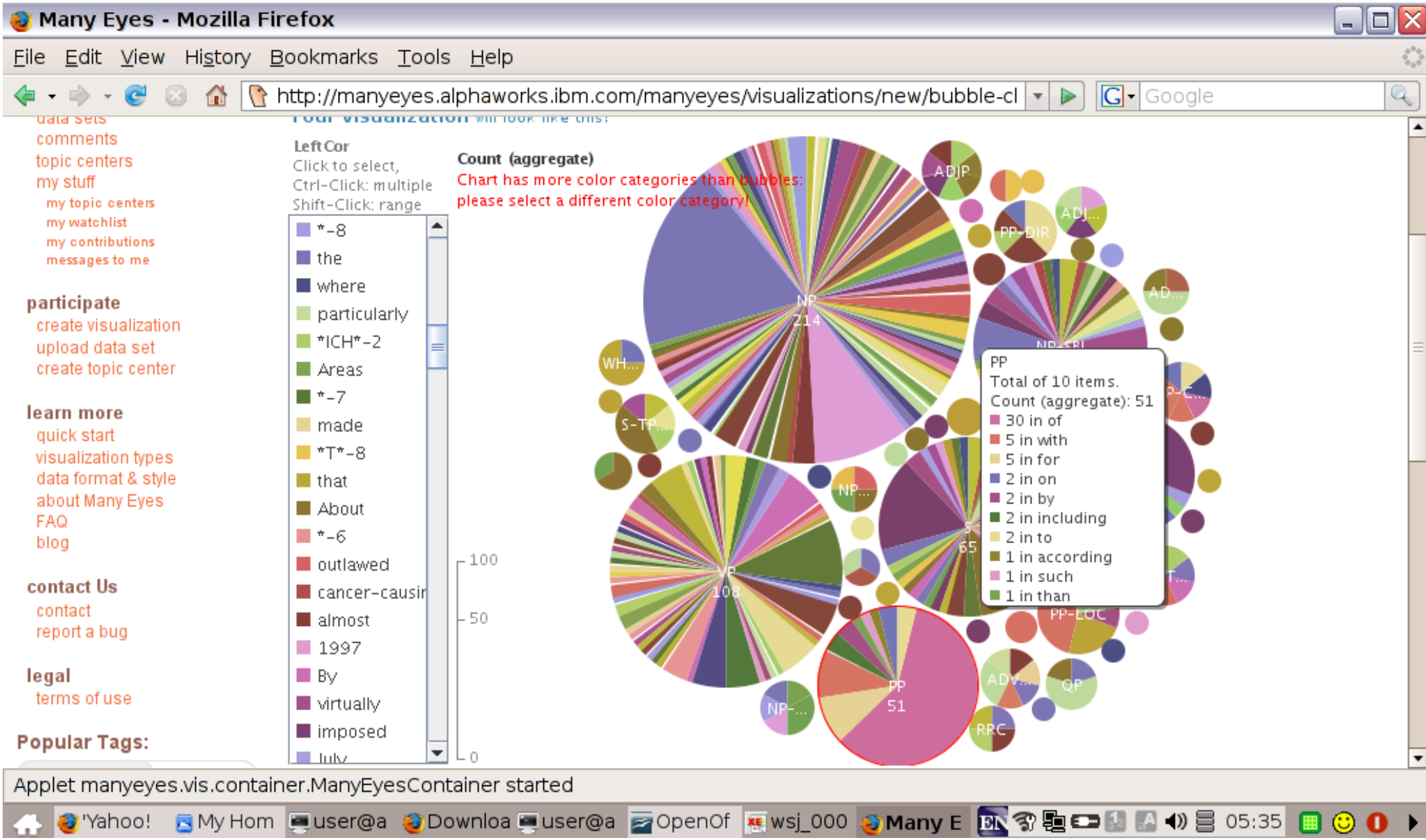
Treemap: parent-child dep's, count by colour



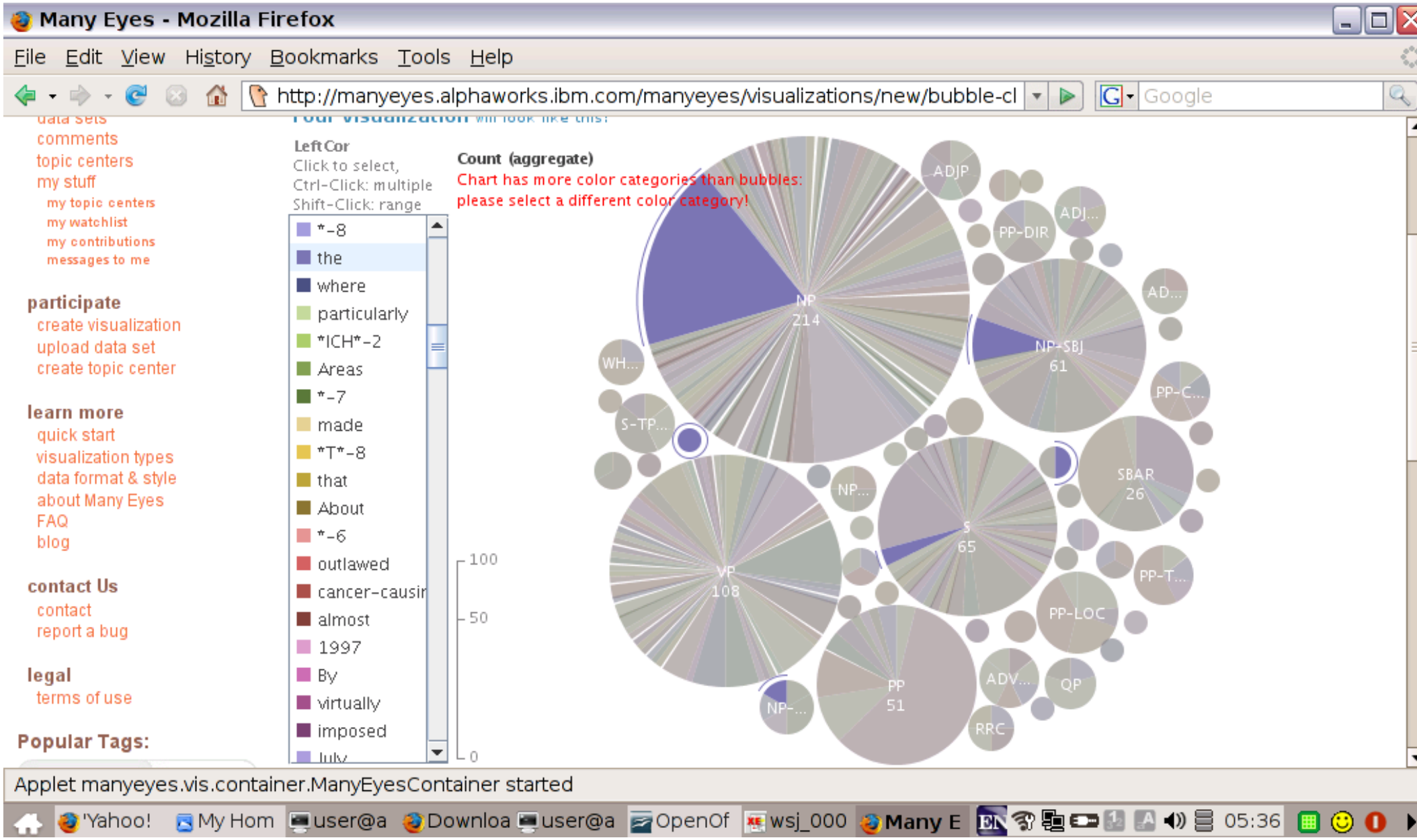
Treemap: parent-child dep's, count by size



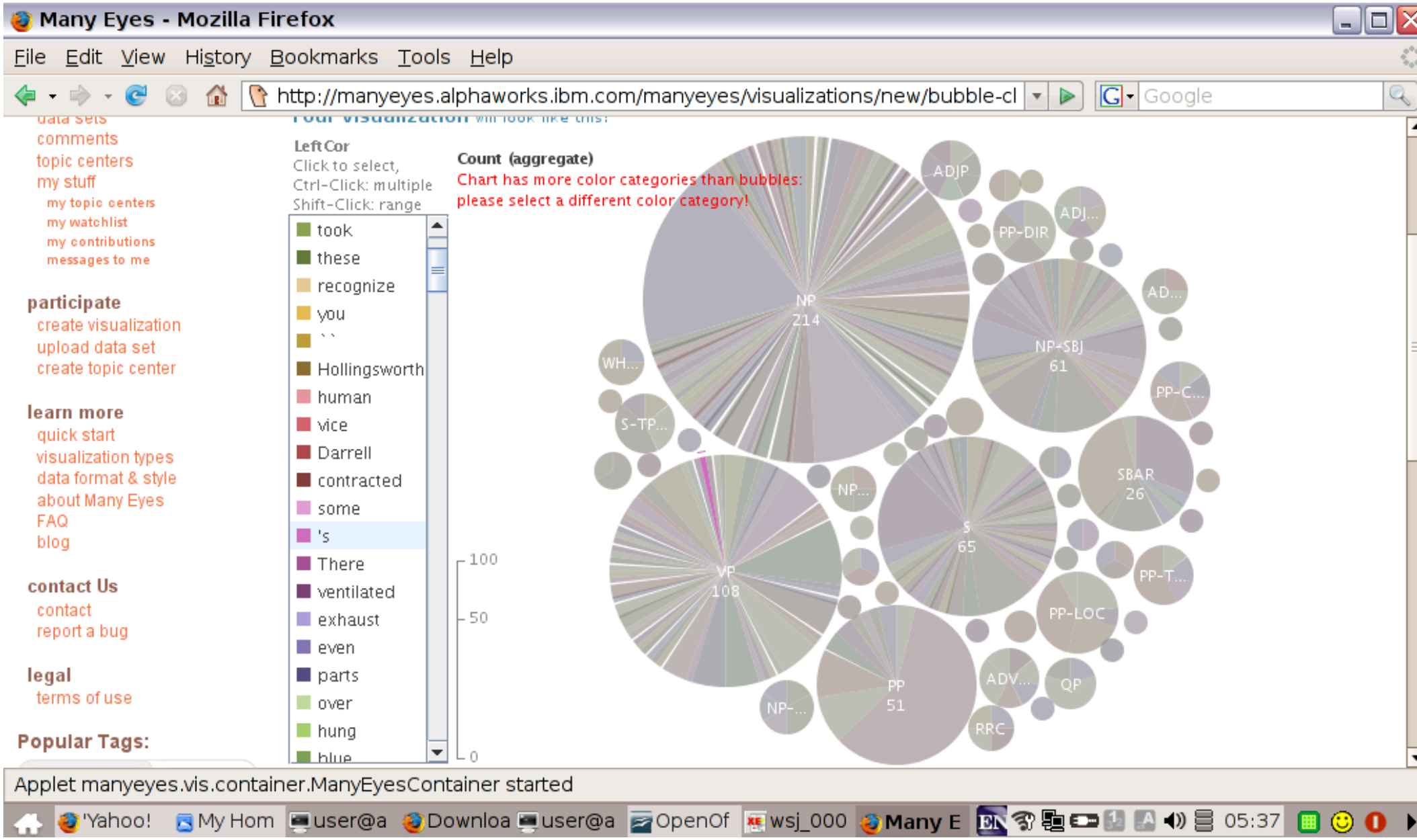
Bubble chart: parent-child dependencies



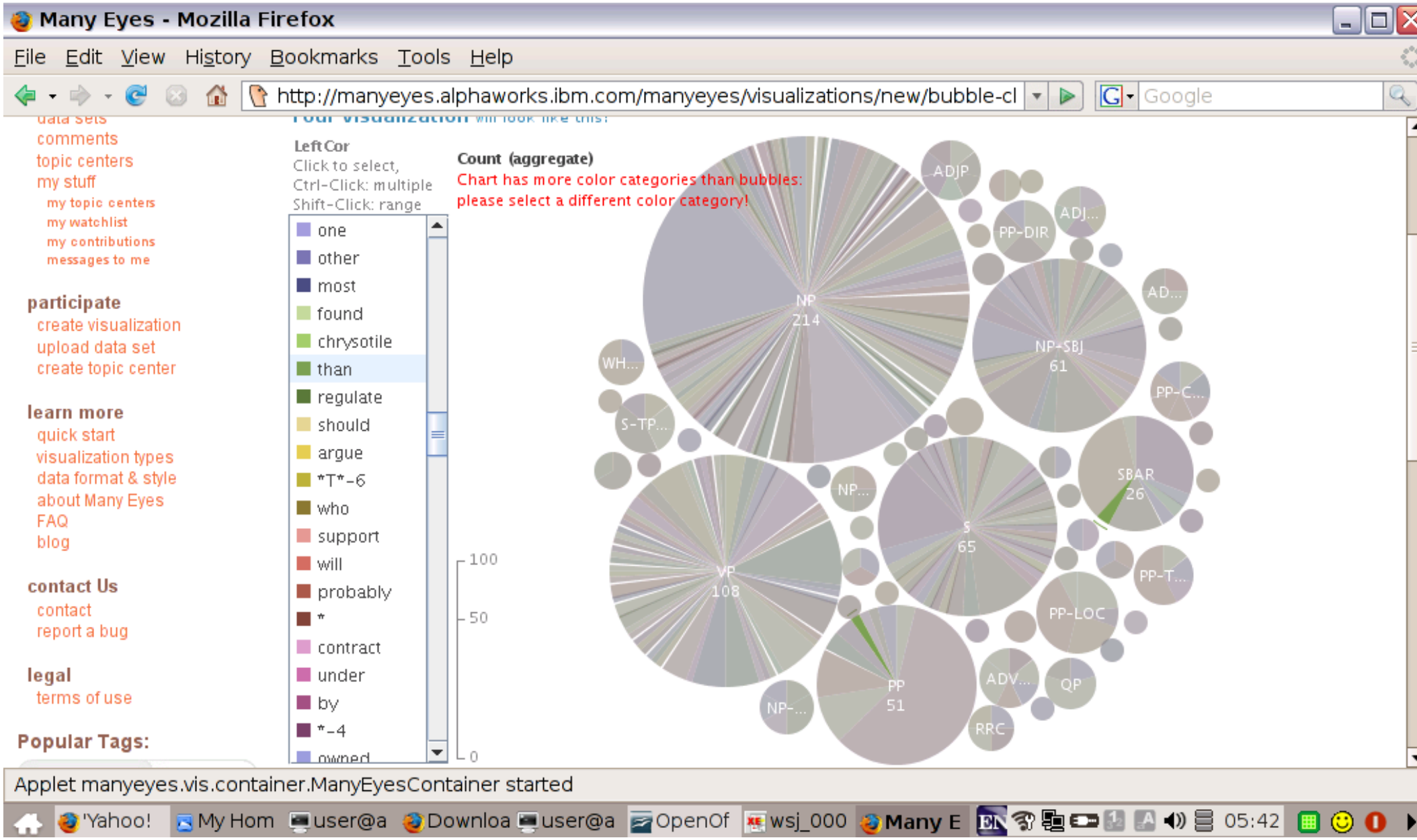
Bubble chart: left-corner dependencies



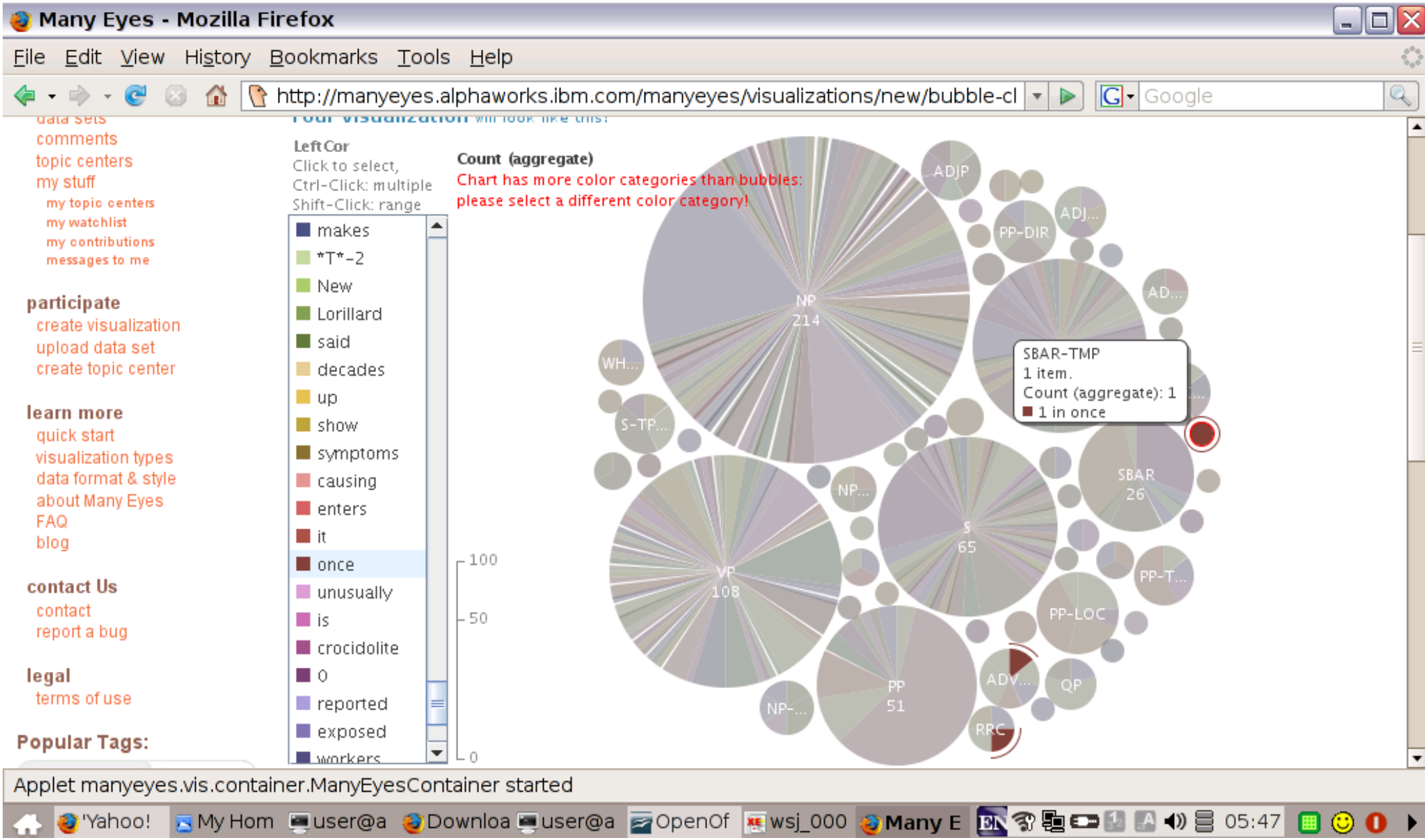
Bubble chart: left-corner dependencies for *the*



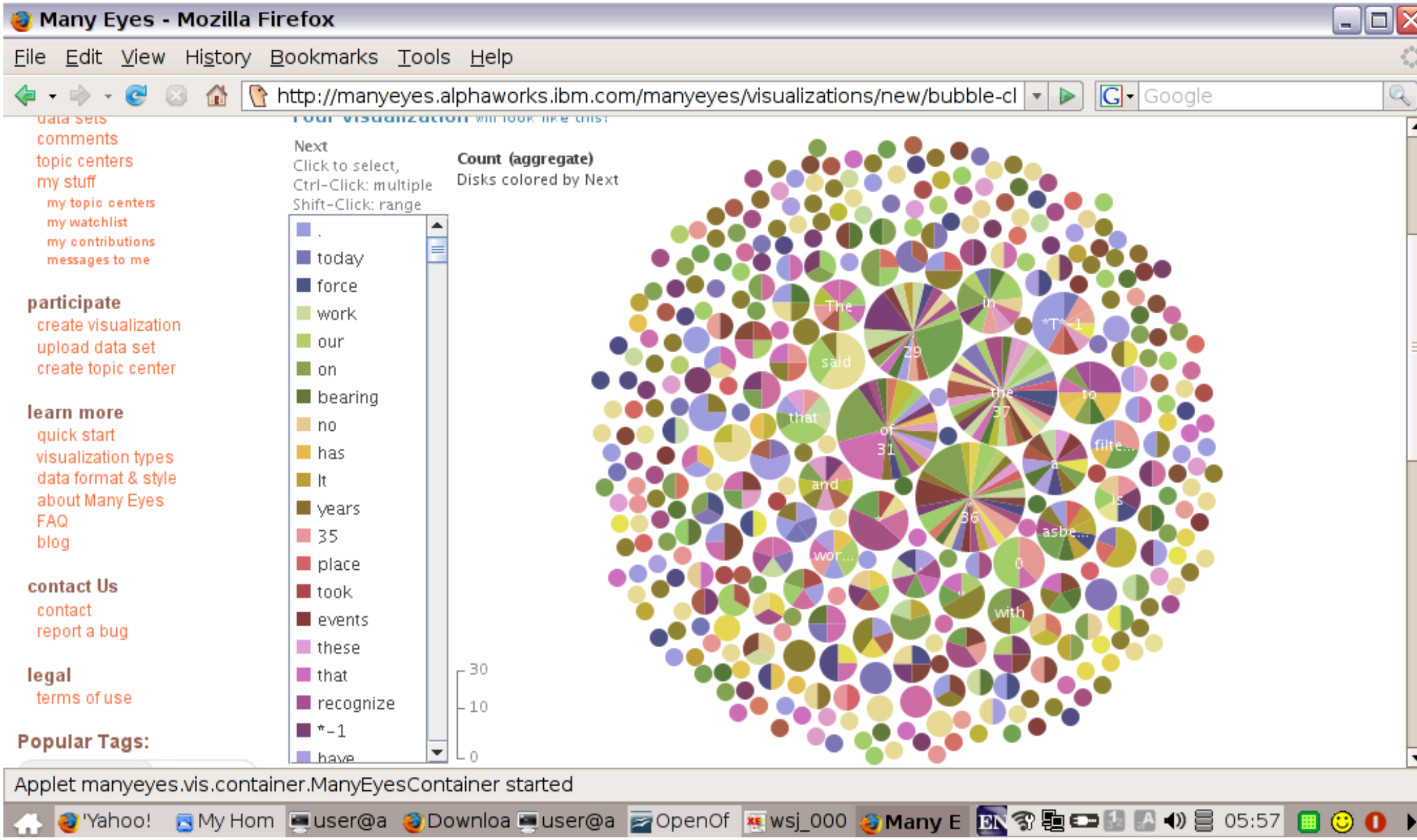
Bubble chart: left-corner dependencies for 's



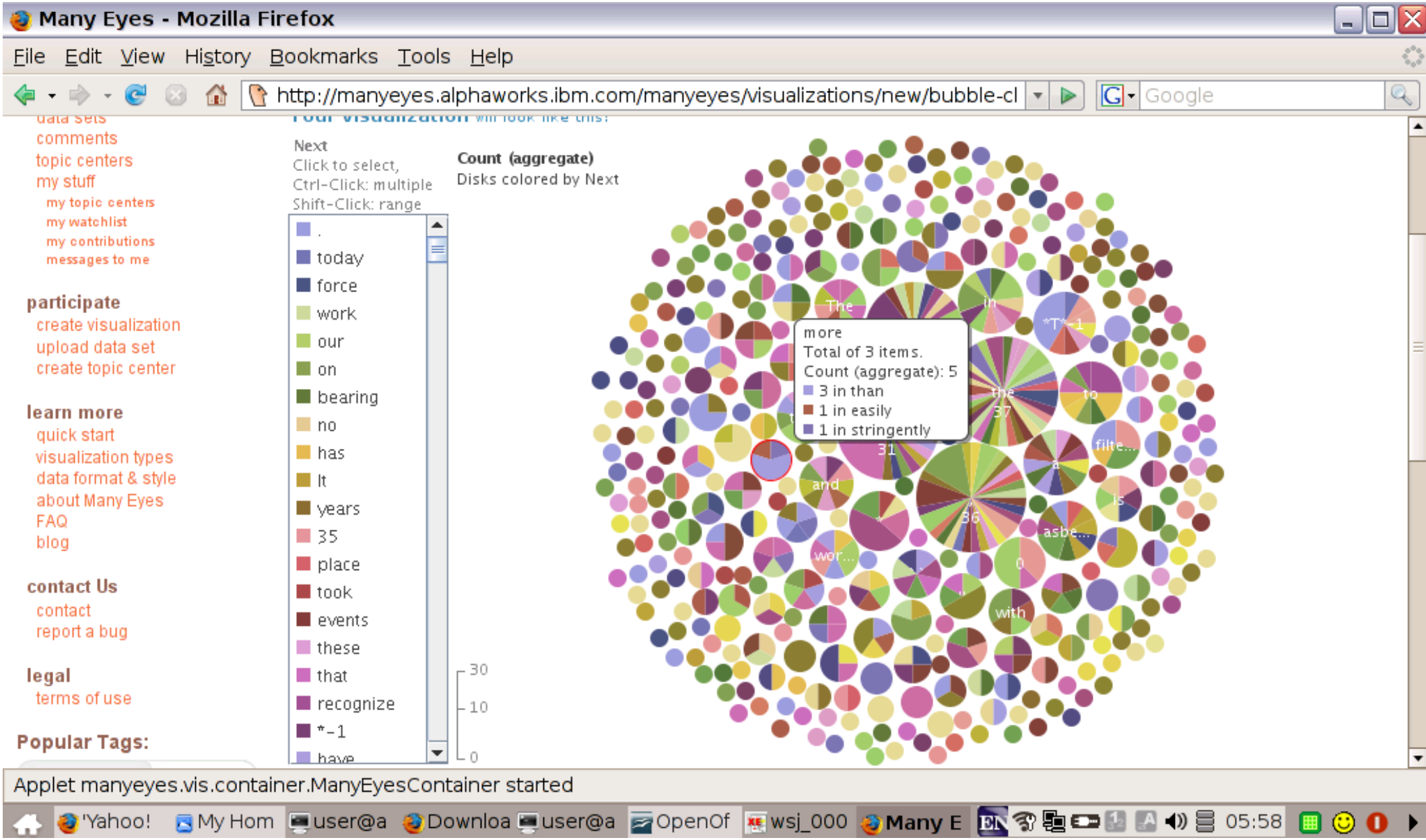
Bubble chart: left-corner dependencies for *than*



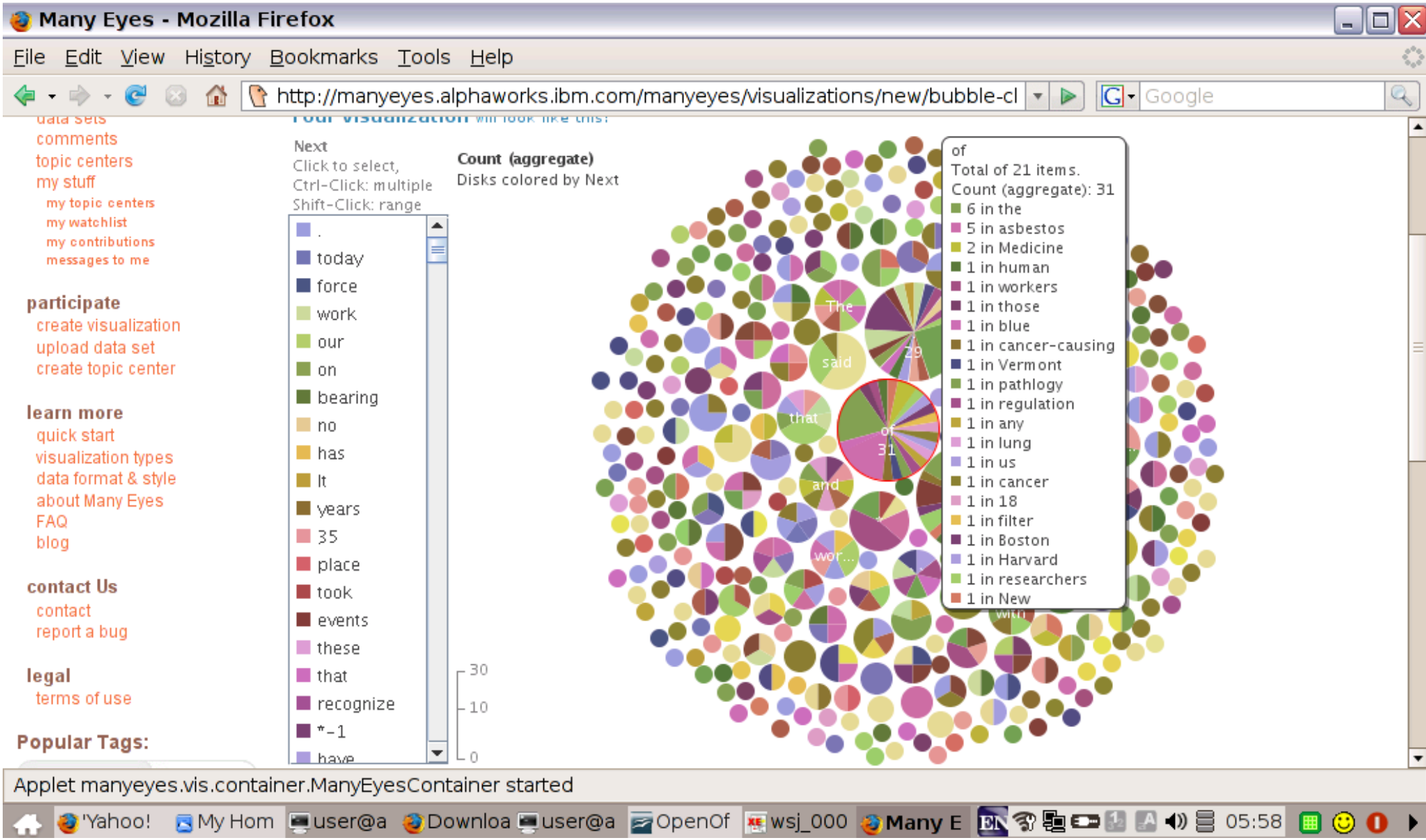
Bubble chart: left-corner dependencies for *once*



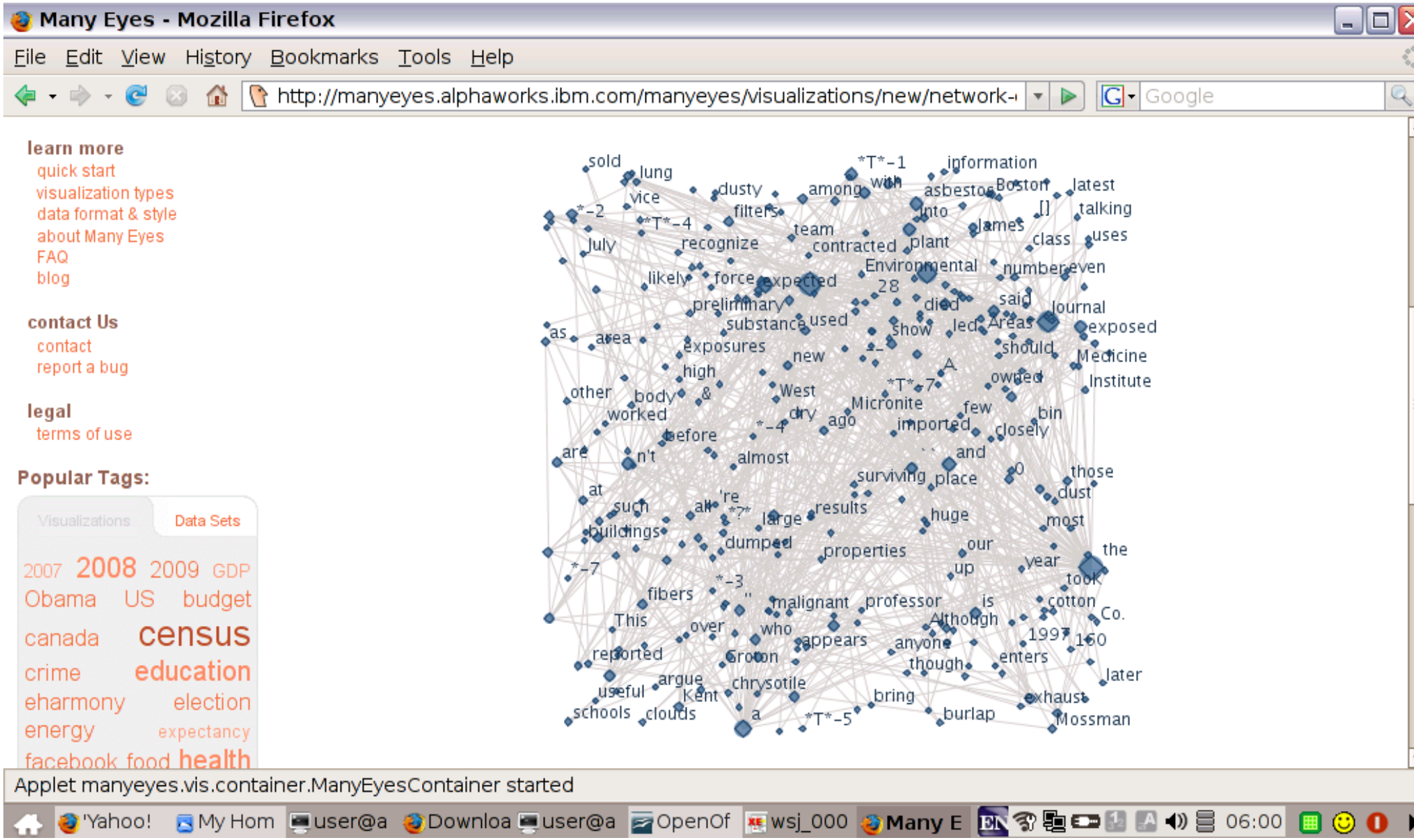
Bubble chart: bi-grams



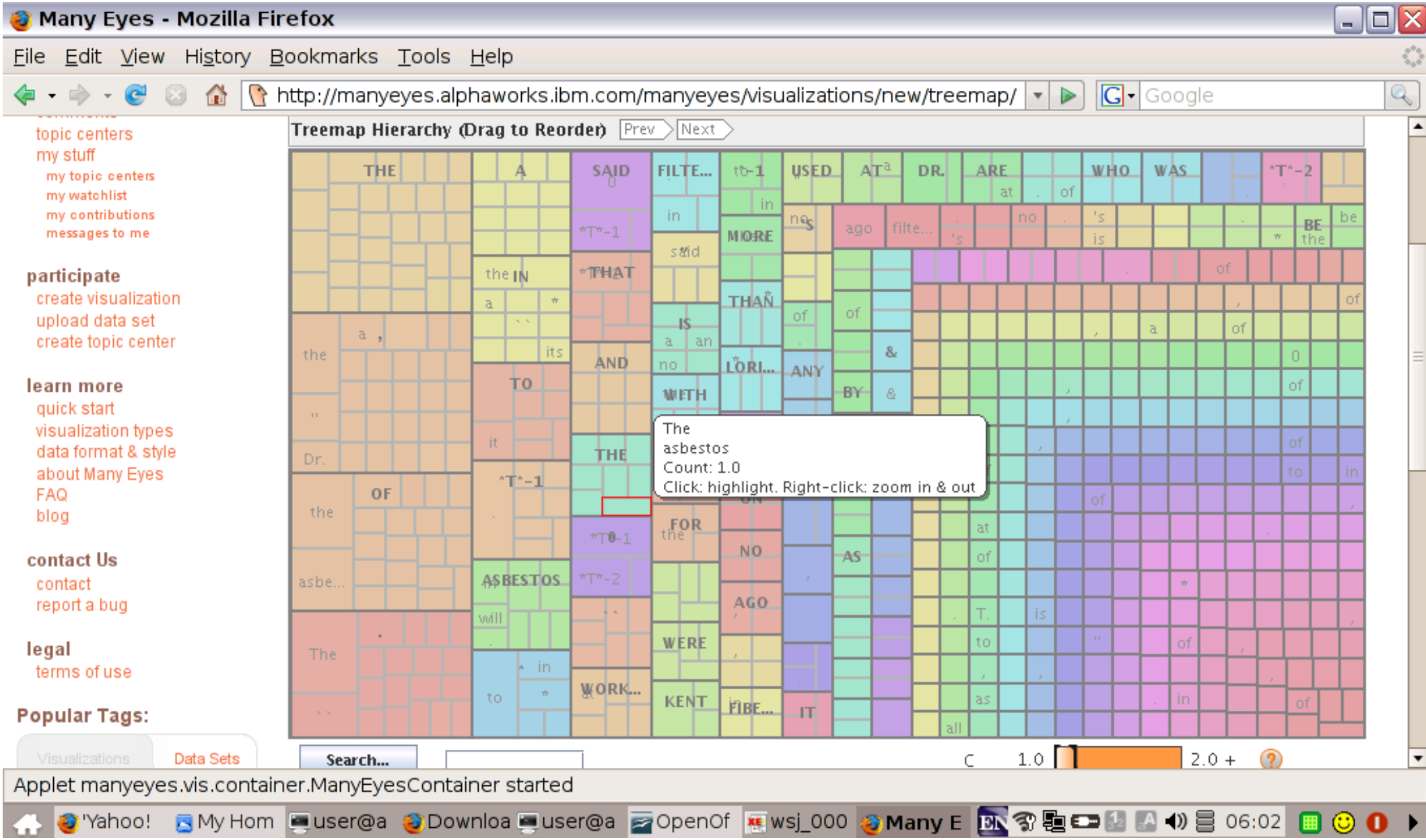
Bubble chart: bi-grams from *more*



Bubble chart: bi-grams from *of*



Network graph: bi-grams



Treemap: bi-grams