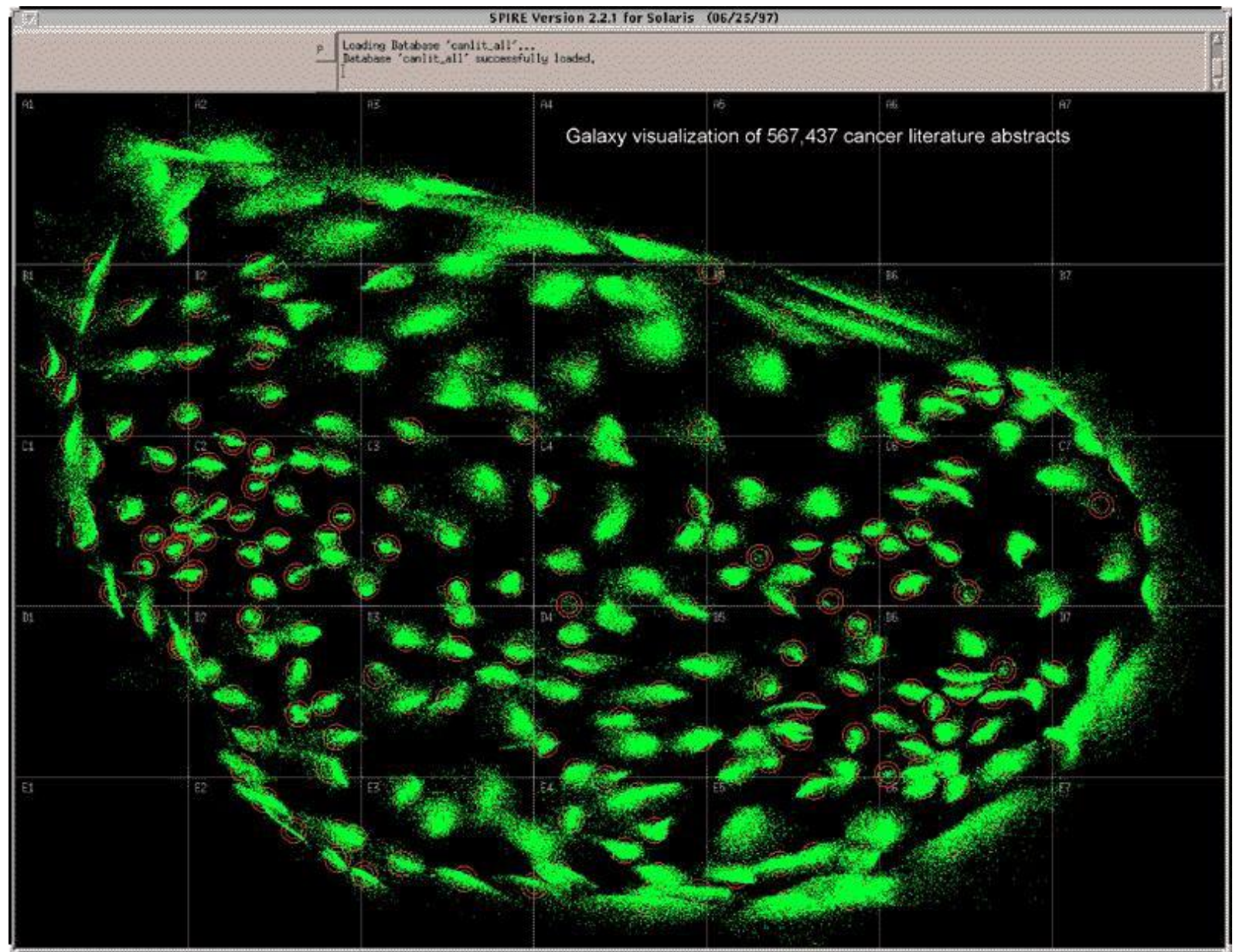
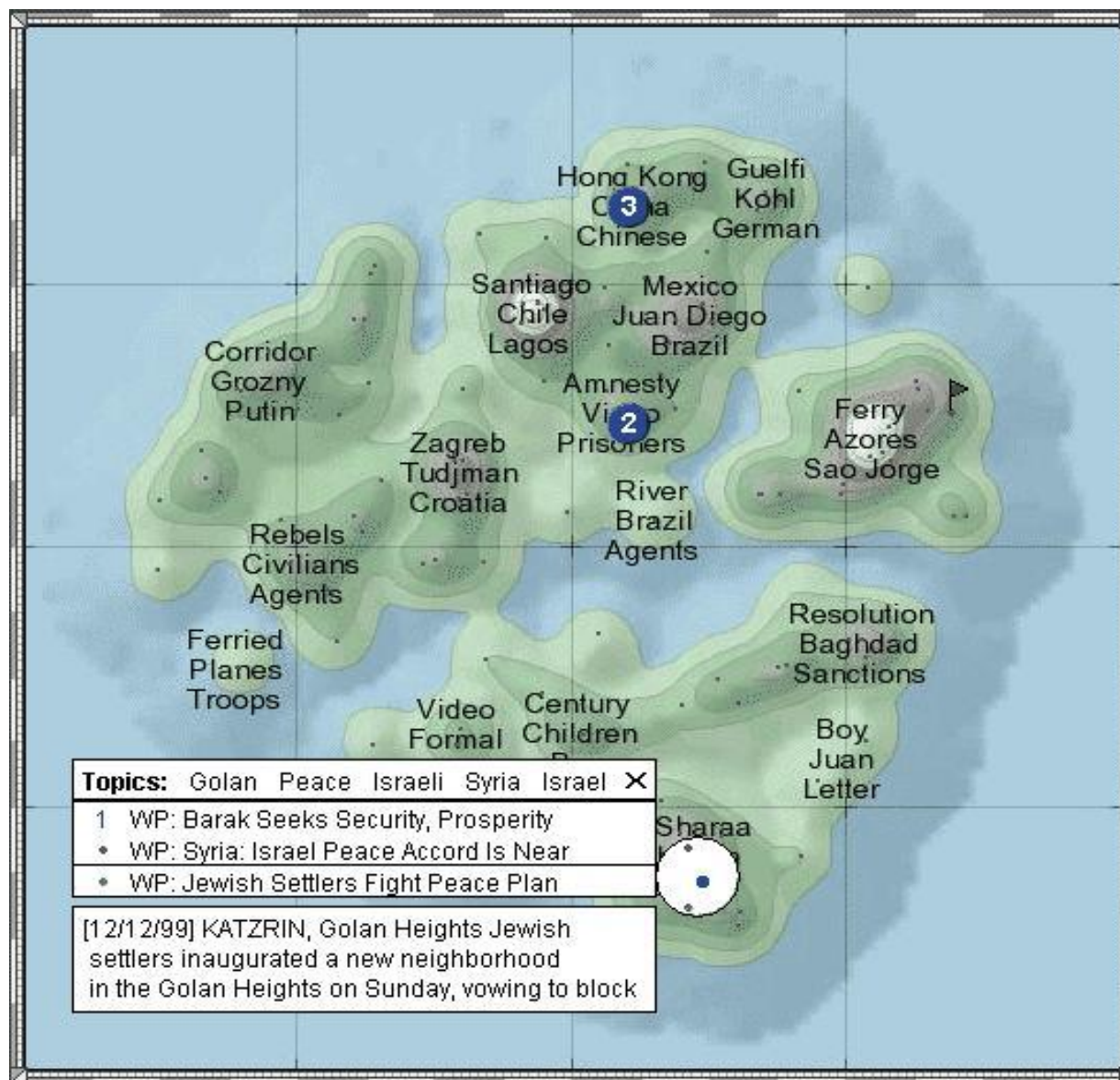
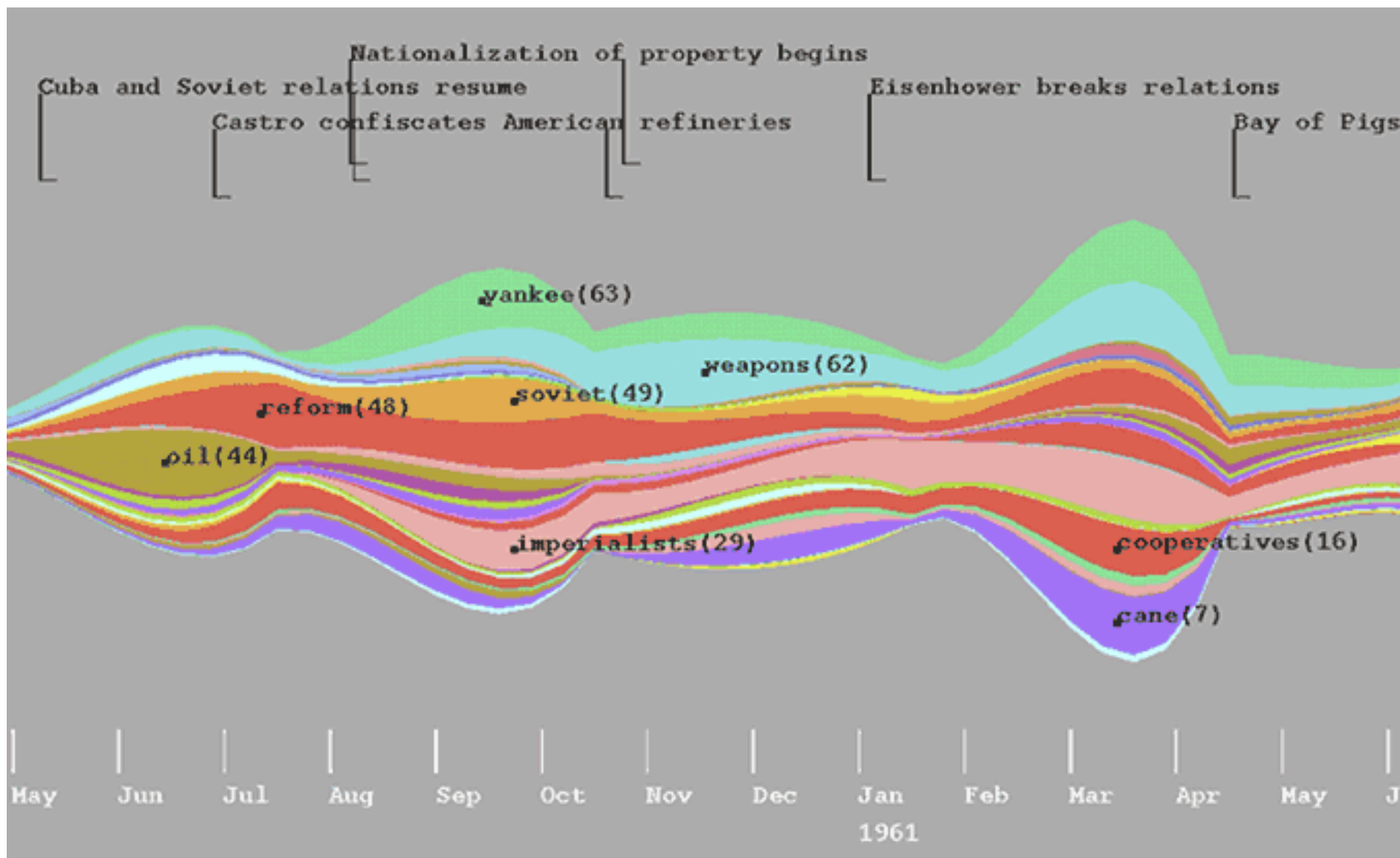




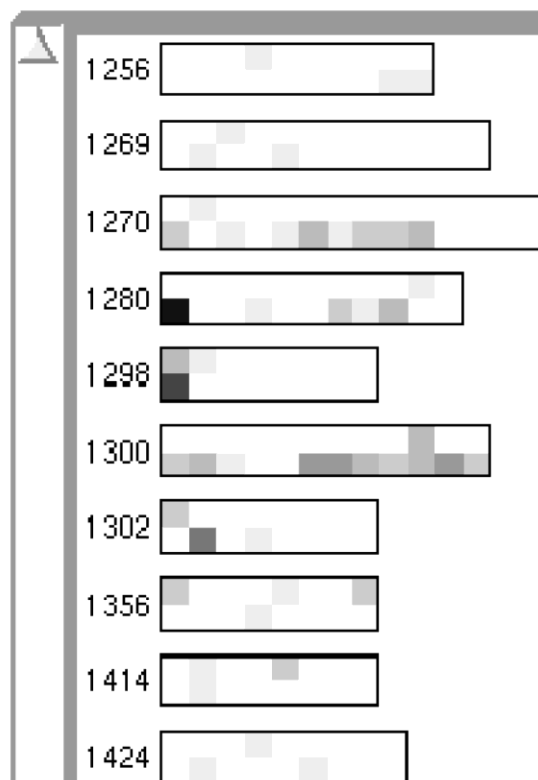




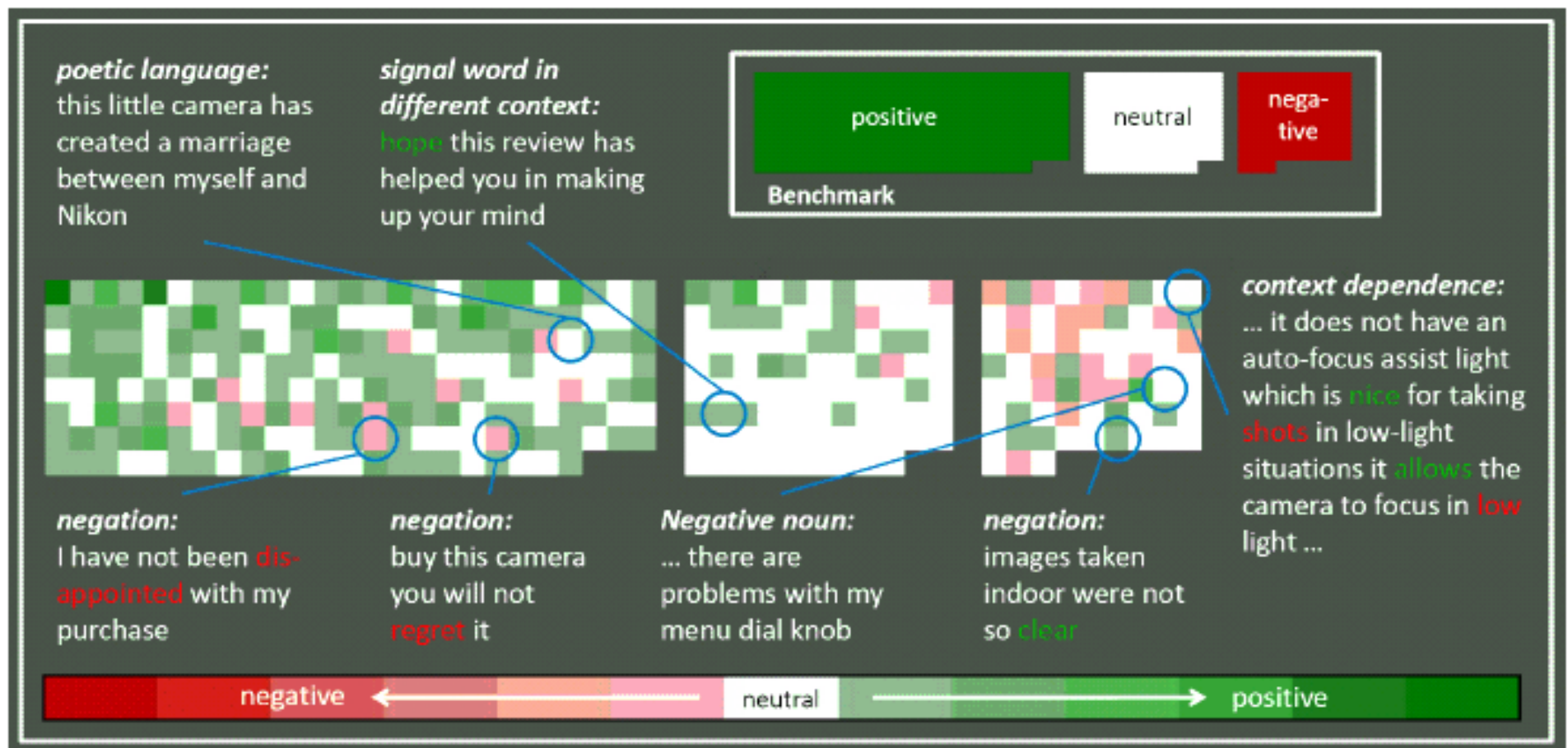
Term Set 1: law legal attorney lawsuit

Term Set 2: network lan

TileBars

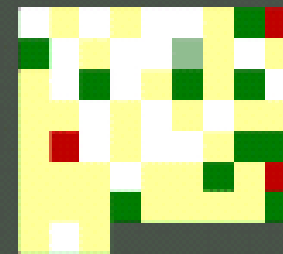
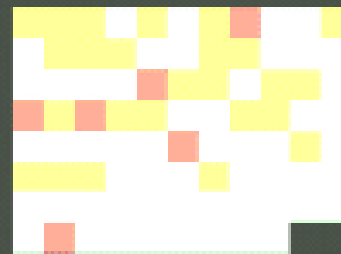
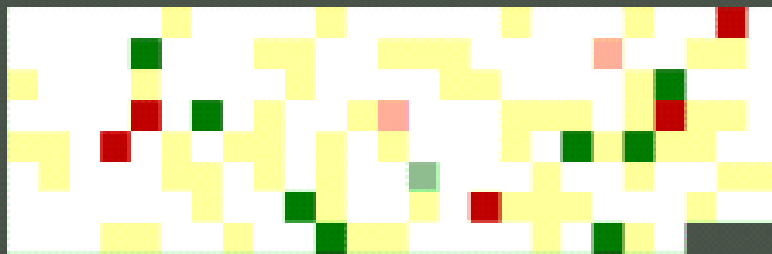


Regression testing handling ha
Toll fraud includes related articl
In conversation Teleglobe Can:
Deregulation indicates a health
The last word letters to the edi
What's wrong with network lice
Letters letter to the editor
Protecting information now vita
Letters O
Loose LIPS sink ships logical ir

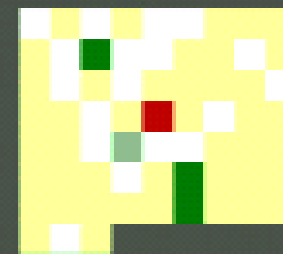
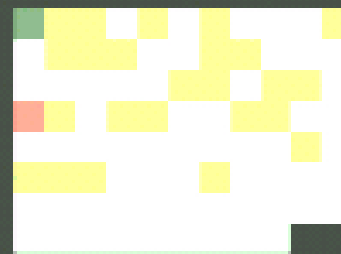
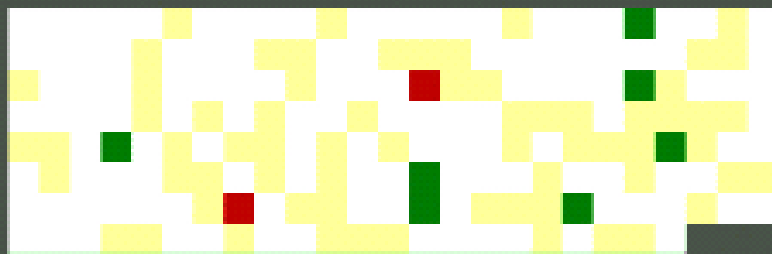


Positive and negative words in Oelke et al. '08 viz.

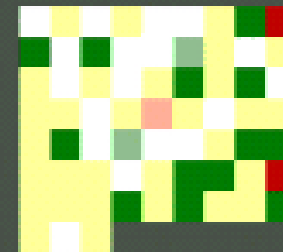
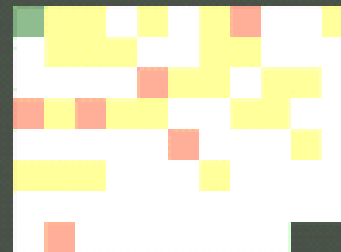
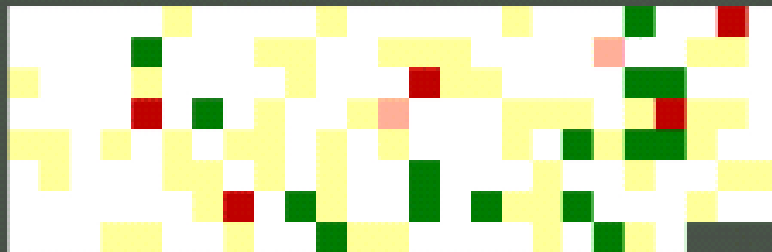
Changes when nouns are included



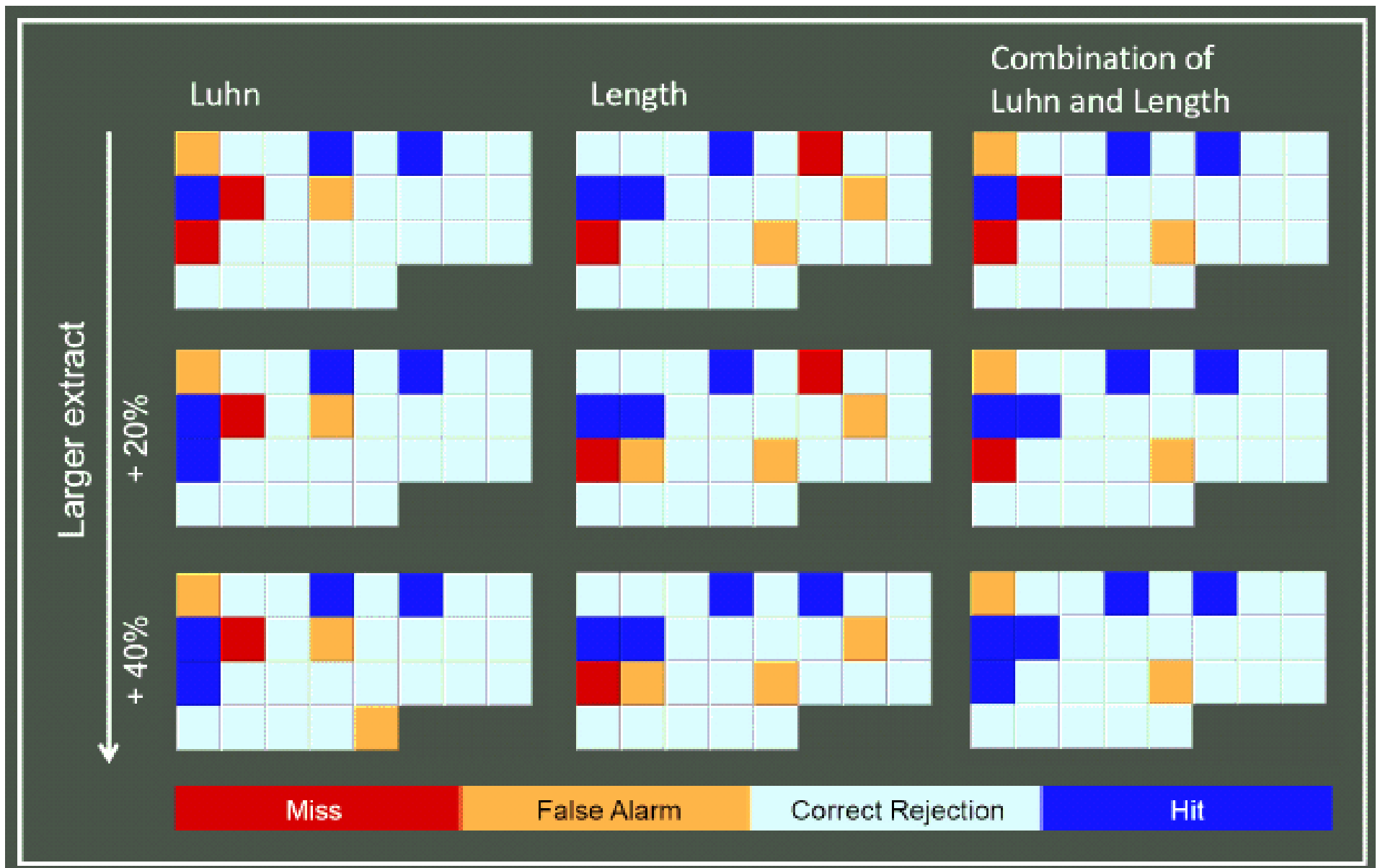
Changes when negation is taken into account



Changes when nouns are included and negation is taken into account



Diff'ing features in Oelke et al. '08 viz.



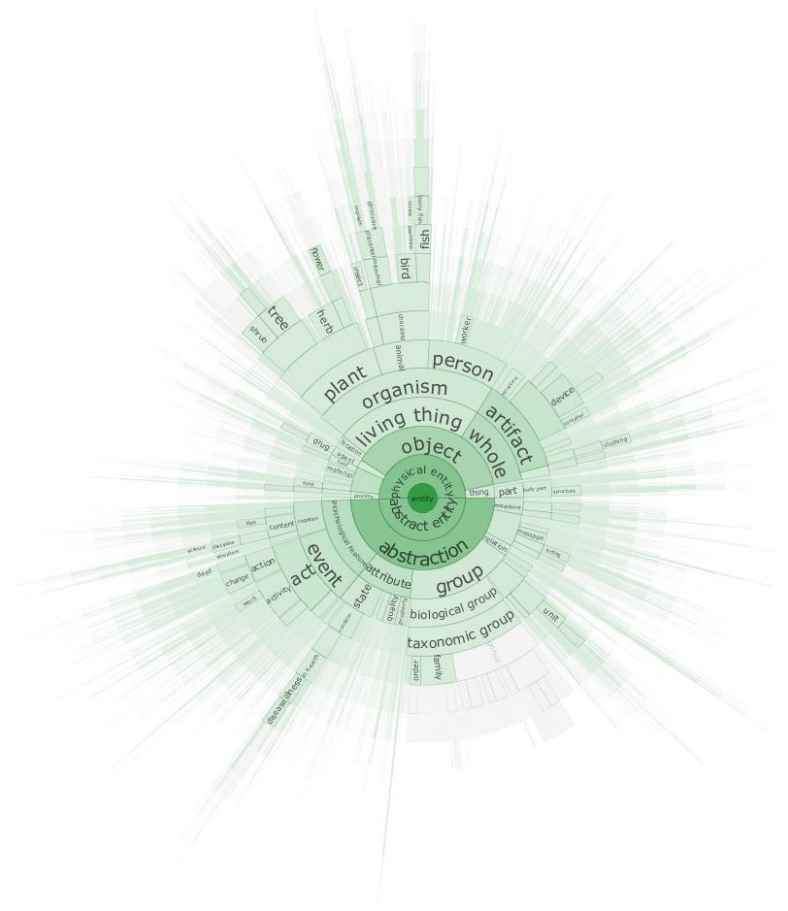
Feature values in a text summarizer - Oelke et al. '08 viz.

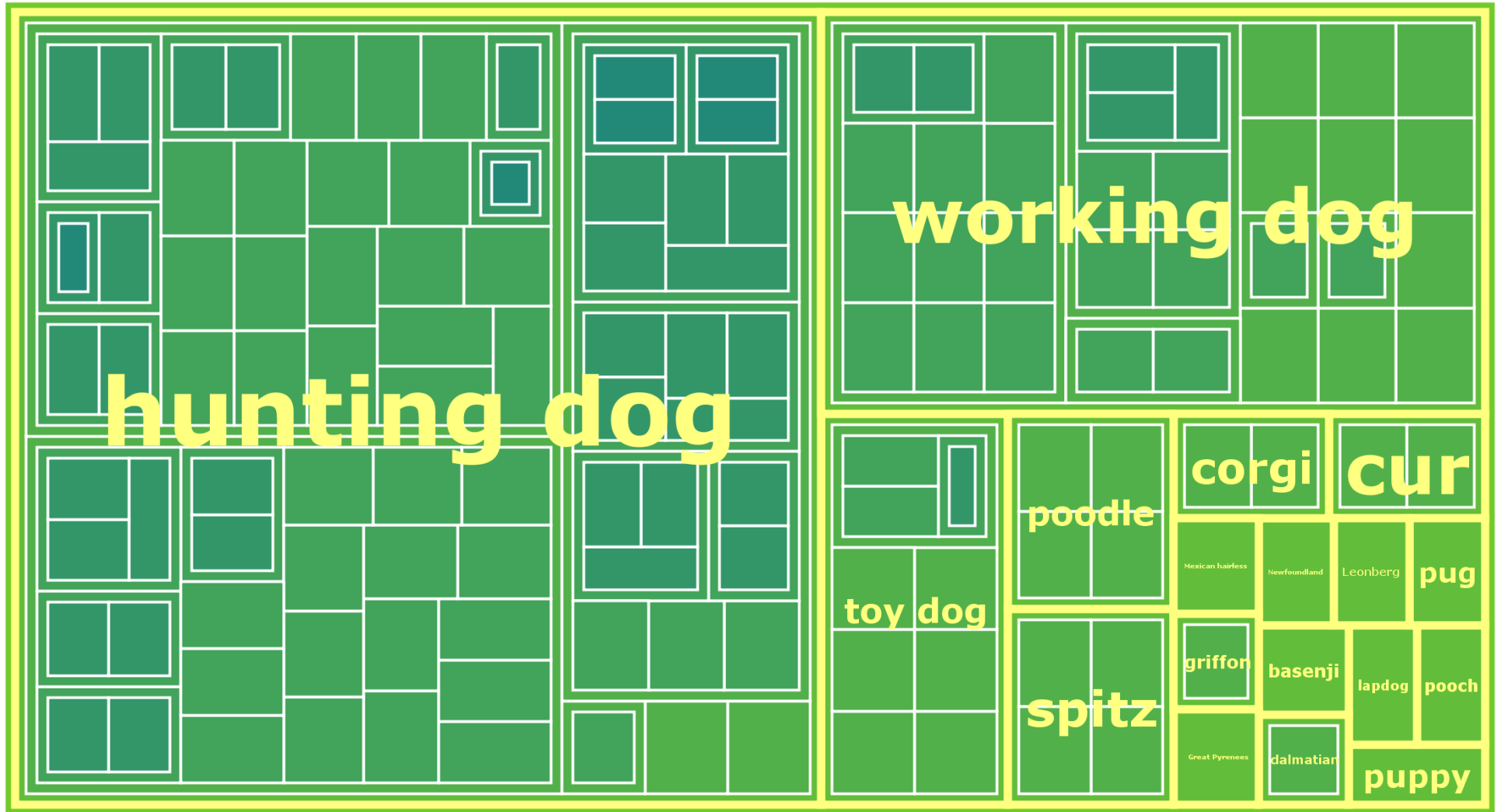


games → game
taken → take

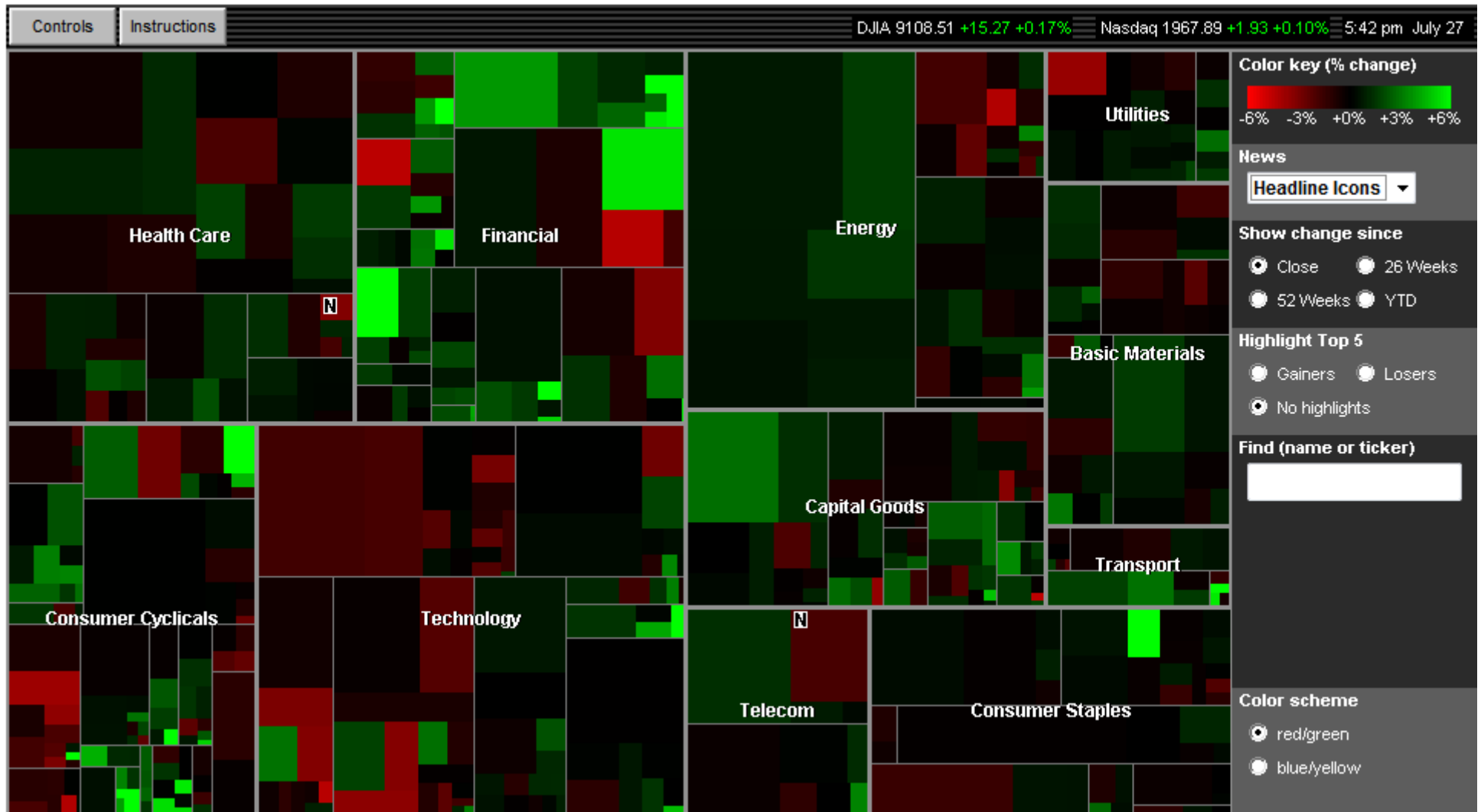
absolute,noun,10
chair,noun,2
moment,noun,11
game,noun,30
reality,noun,3
take,verb,13
represent,verb,17
...

game IS activity
chair IS furniture

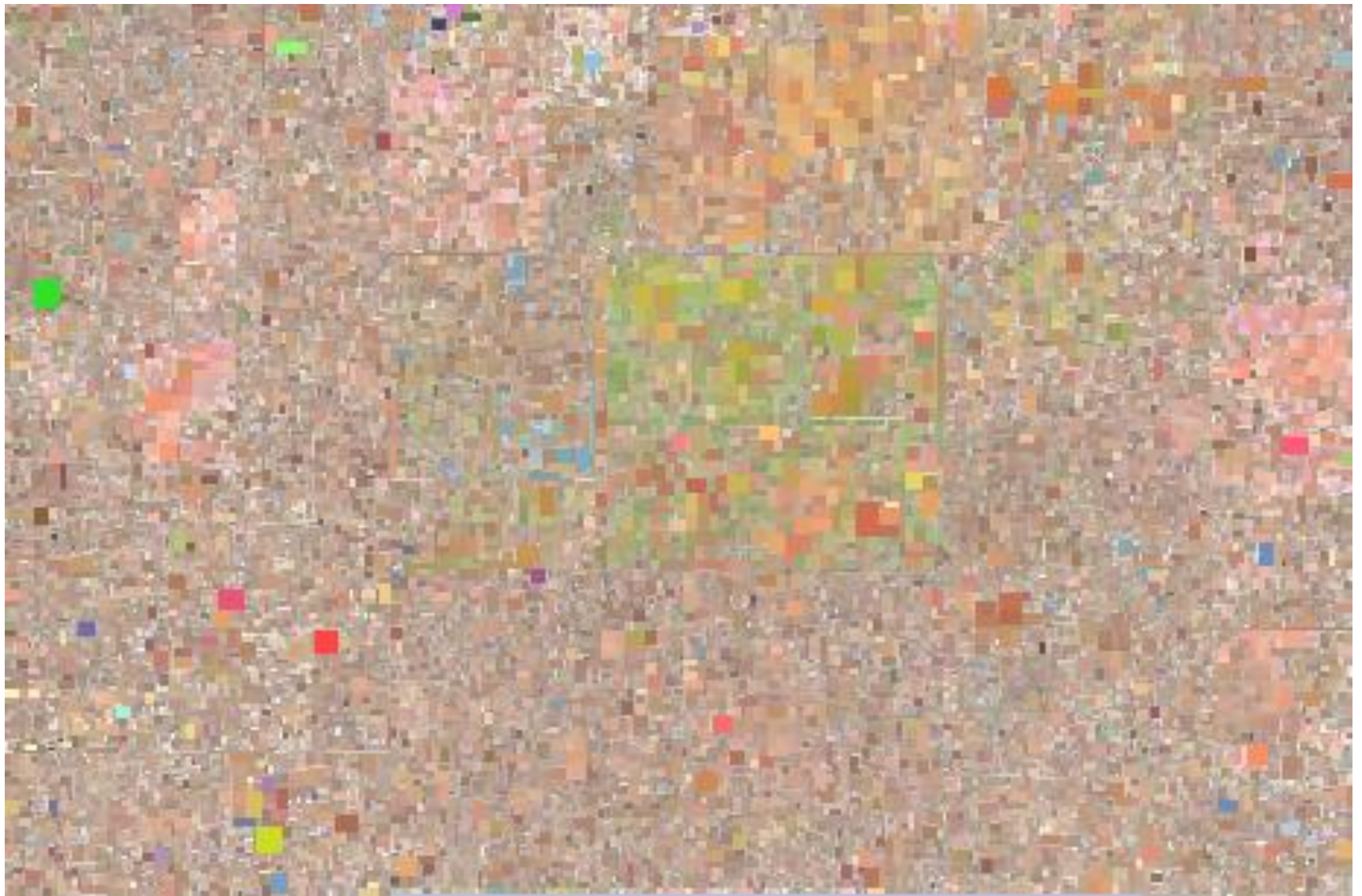




Tree map viz (Schneidermann and Johnson '91)
of WordNet



Tree map viz (Schneidermann and Johnson '91)
of stock market [www.smartmoney.com]



DerivTool 0.6
New Corpus
Corpus=f-test.50 (50 lines) 20
Go to line
Load
Save
Print
Print to File
Redraw
Close

Derivation Tree
Freeform Notes

English: the gunman was killed by the police .

Click, Ctrl-click, or drag above to select. For adding rules, please select a contiguous span of top level nodes.
Modify
Delete

Phrase-based MT: gunman killed by police .

Searching
Manual
Selecting Template

Lines = 100
Redraw
Click below to add rule. (Red=no rules found, Blue=used by AT translation, Green=default translation, Purple=Red+Blue)

"枪手"	"被"	"警方"	"击毙"	"."
gunmen	being	police	killed	.
by gunmen	were	traffic police	kill	,
gunman	been	police officers	yugoslav guards responded and killed	
assailants	are being	police force	killing	
shooters	had been	police are		
	have been	police have		
	has been	a police		
		the police		
	by police			
	was			
	are			
	be			
	by			

no rules are available for this phrase (red)

this phrase was chosen by the phrase-based decoder (blue)

no rules are available for this phrase and it was chosen by the phrase-based decoder (purple)

The Chinese Room

Edit Style

Send: Prev Next Doc: Prev Next Pop

Show Parse Show Alignments

Select Translation Edit Examples N-Best Translations

Two years ago this month

Guangdong

new high tech

this year

...years ago

Guangdong province, China

technology, technique, skill

today, modern, present, current, this, now

year

before, in front, a former, previous, a front

both, two, ounce, a few, teal

moon, month

wide, numerous, to east

high, tall

new, newly

skill

to reproduce, to pr

Results for phrase: 今年前

子 今年 前 年 前

...and Security Affairs earlier this year from the Megawati ...

8.679999

今年 前 3

...economy for the first three months of this year grew only ...

8.04

今年 前 6

...year within the first 6 months, ...

8.200001

今年 前 7

...Qi Zhongxi -RRB- In the first seven months of this year, the growth ...

7.72

在 今年 年底 前 彻底

...eliminate globally by the end of the year the disease that ...

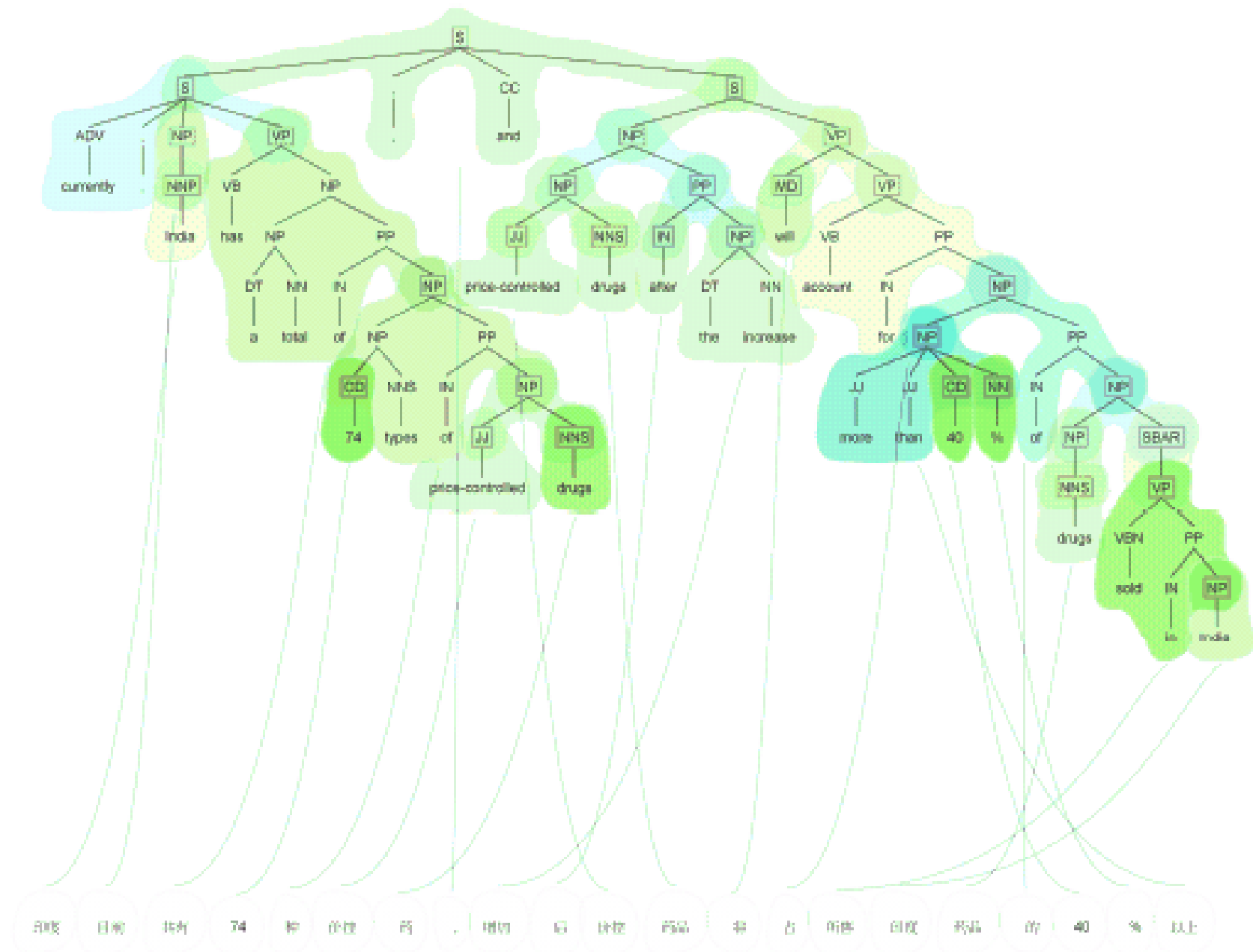
7.36

沙赫巴兹 谢里夫 今年 7月 接替 他的 哥哥、前 总理

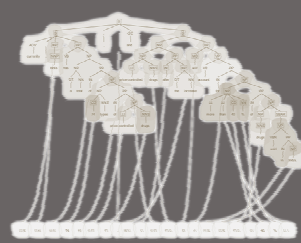
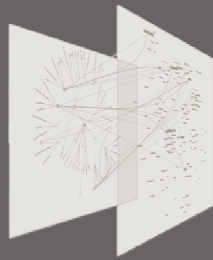
Shahbaz Sharif replaced his elder brother, former Premier Nawaz Sharif, as the leader of Pakistan Muslim League in July this year. 80.49

100%

“Chinese Room” MT viz (Albrecht et al. 09)



Bubble Graphs MT viz (Collins et al. '09)



Interactive Visualization for Computational Linguistics

ESSLII 2009

2

Information Visualization Basics

Perception

3

Pre-attentive processing

4

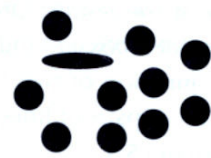
orientation



curved/straight



shape



shape



size



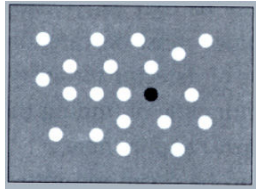
number



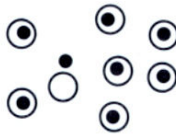
Ware

Pre-attentive processing

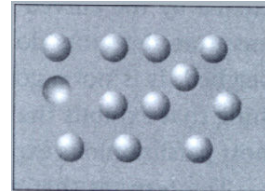
value



enclosure



convex/concave



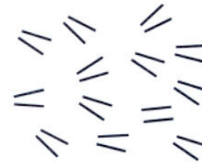
addition



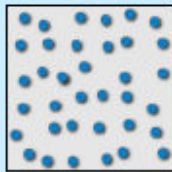
Juncture ??



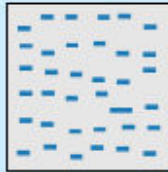
Parallelism ??



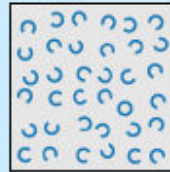
Pre-attentive processing



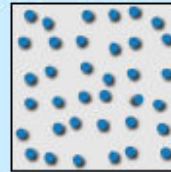
depth



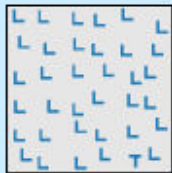
Length/width



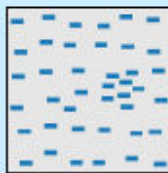
closure



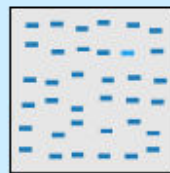
lighting direction



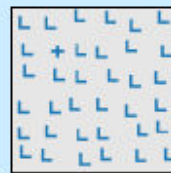
terminators



density



intensity



intersection

<http://www.csc.ncsu.edu/faculty/healey/PP/>

Graphical Excellence

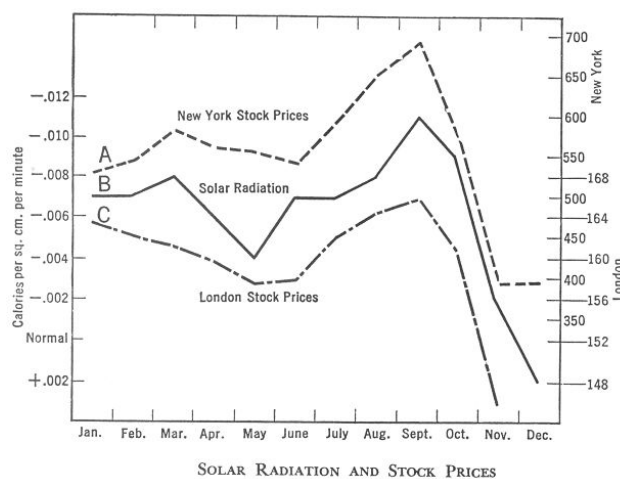
- Principles of **graphical excellence**
 - ▣ Complex ideas presented with clarity, precision, honesty, and efficiency.
 - ▣ Gives viewer the greatest number of **ideas** in the shortest **time**, with the least **ink** in the smallest **space**.
- Graphical excellence often found in **simplicity of design** and **complexity of data**.
 - ▣ multivariate data, simple design

Graphical Excellence

- Fundamental graphic designs include:
 - Data maps: these involve placement of additional information spatially situated on a spatially explicit diagram.
 - Time series: these involve plotting some data as it changes across time.
 - Space-time narrative: plotting changes across both space and time.
 - Relational graphics: designed to show the relationship between two or more data aspects.

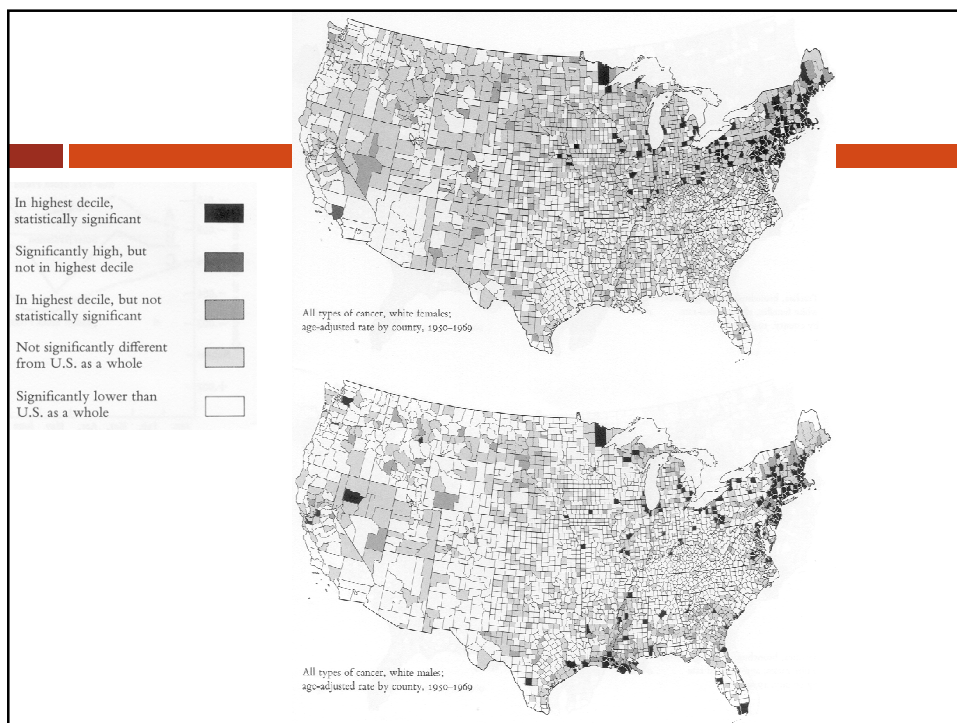
Graphical Excellence

- Start with reasonable data
- A. New York stock prices
- B. Solar radiation inverted,
- C. London stock prices
- For all months 1929



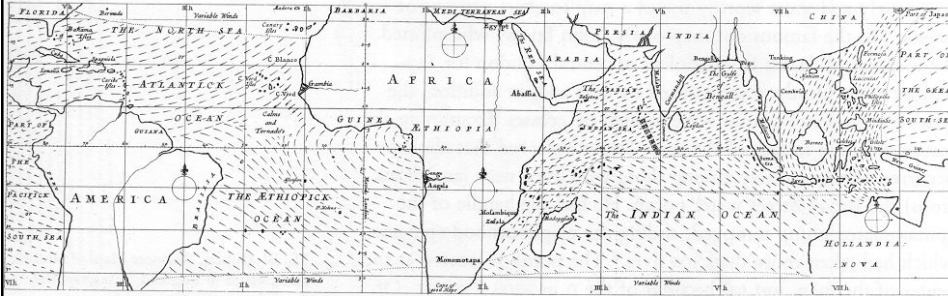
Data Maps

- Age-adjusted death rates by cancer type for USA
- (each some 21,000 numbers)
- Includes overall pattern and county by county detail
 - High death rates in north east and around great lakes
 - Low rates in band down middle
 - Higher rates for men than women in south
 - Hot spots; in Minnesota, Iowa, Nebraska, along the Missouri River
- Atlas of Cancer Mortality in the U.S., 1950-94 (Book)
- <http://www3.cancer.gov/atlasplus/index.html>



Data Maps

- 1st data maps
- Edmond Halley's 1686 map shows trade winds and monsoons

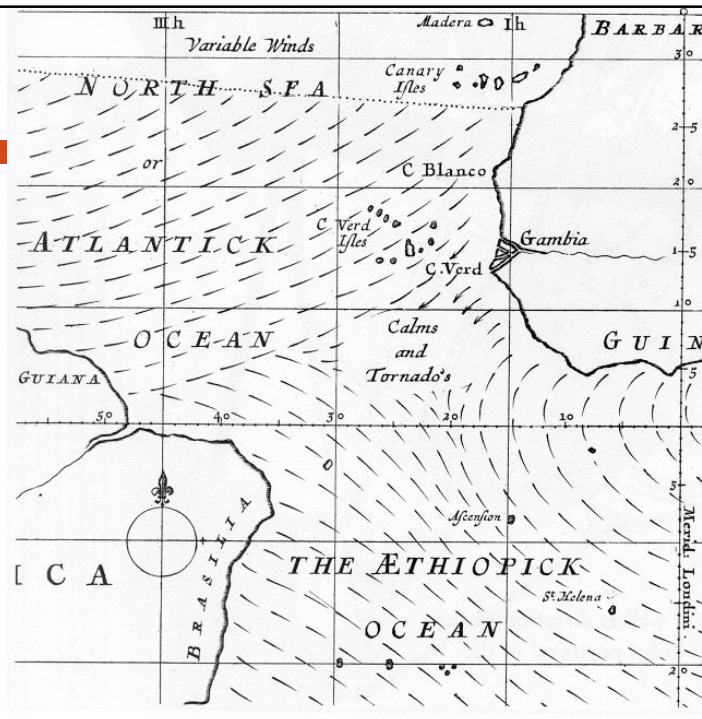


Norman J. W. Thrower, "Edmond Halley as a thematic Geo-Cartographer," *Annals of the association of American Geographers*, 59 (Dec. 1969), p 652-676 (Tufte, 1983, p.23)

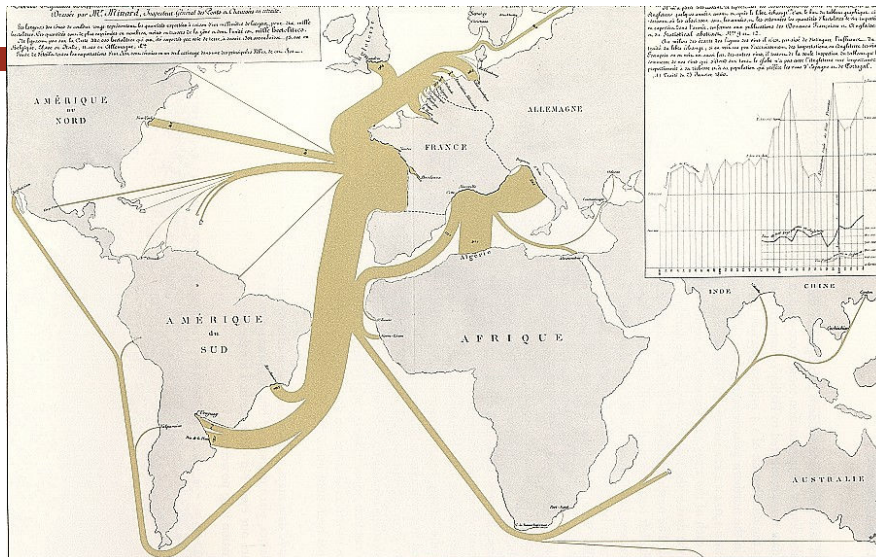
Data Map

- 1686
- Detail of trade winds map shows encoding – "the sharp end of each little stroke pointing out ... from whence the wind continually comes"

Edmond Halley, "An Historical Account of the Trade Winds, and Monsoons, Observable in the Seas Between and Near the Tropicks; With an Attempt to Assign the Physical Cause of Said Winds," *Philosophical Transactions*, 183 (1686), p. 153-168 (Tufte, 1983, p.23)



Data map: 1864 Exports of French Wine

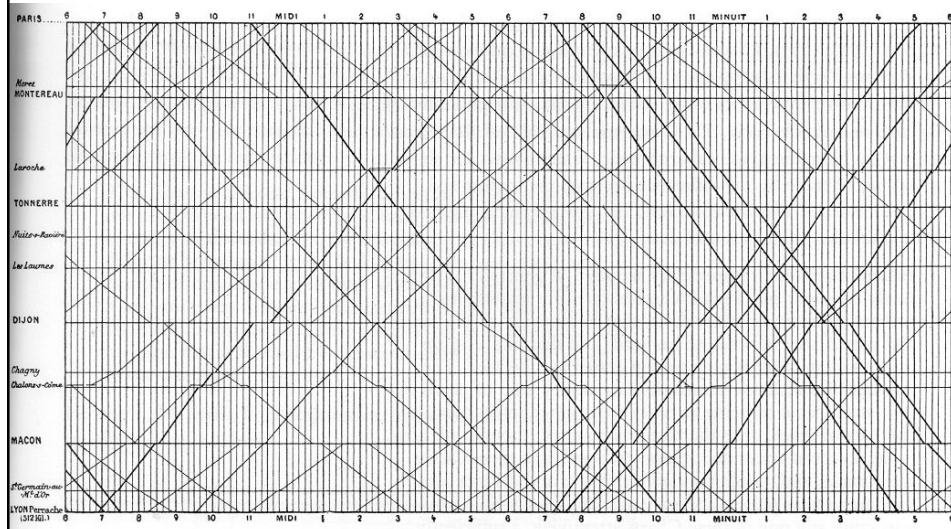


E. Tufte "Visual Display of Quantitative Information" p 25,

Time Series

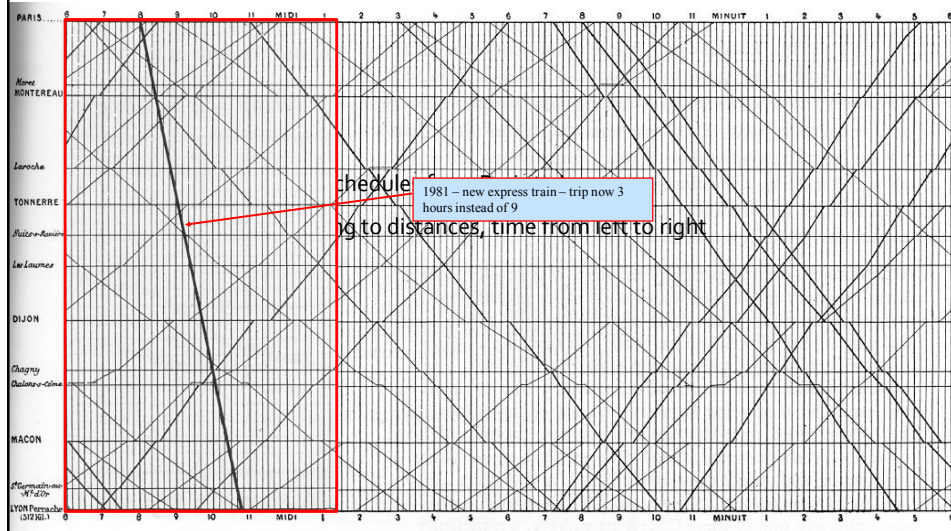
E.J. Marey, "La Methode Graphique," (Paris 1885), p.20. This method is attributed to the french engineer, Irby (Tufte, 1883, p.31)

- E. J. Marey. 1885. Train schedules from Paris to Lyon
- Stations spaced according to distances, time from left to right



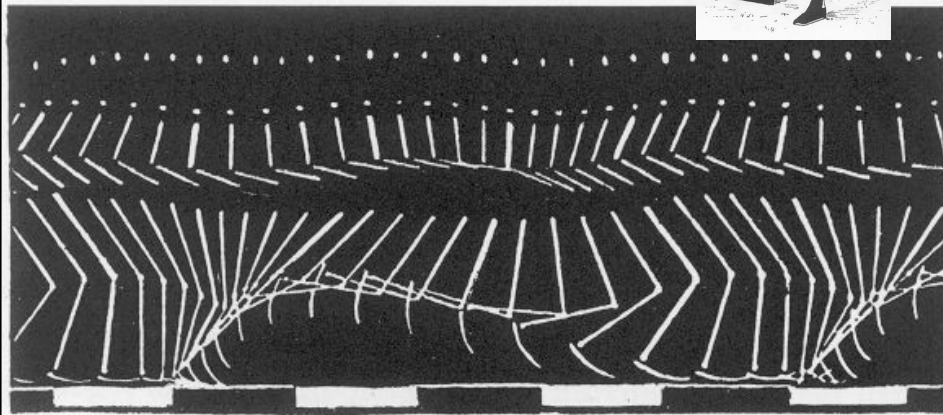
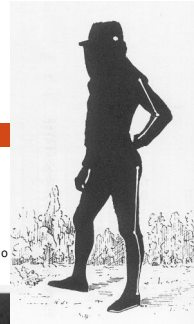
Time Series

E.J. Marey, "La Methode Graphique," (Paris 1885), p.20. This method is attributed to the french engineer, Irby (Tufte, 1883, p.31)



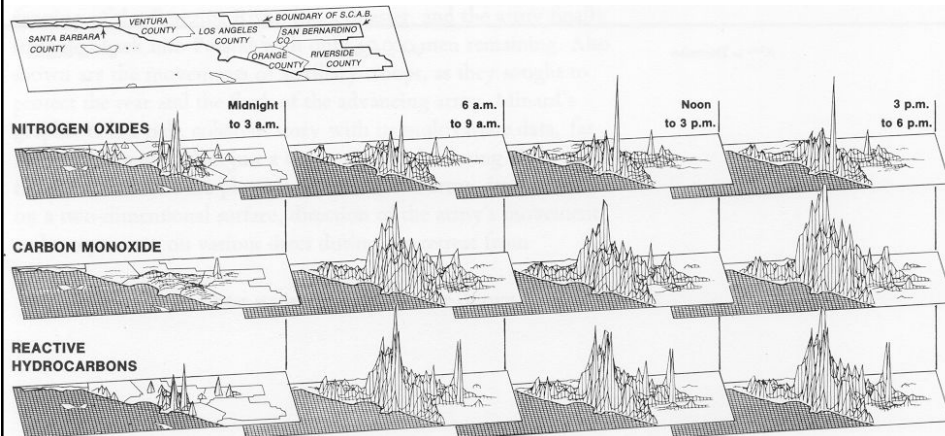
Time Series

E.J. Marey, "Movement," (London 1895), p.60 and 61. (Tufte, 1883, p.35-36)



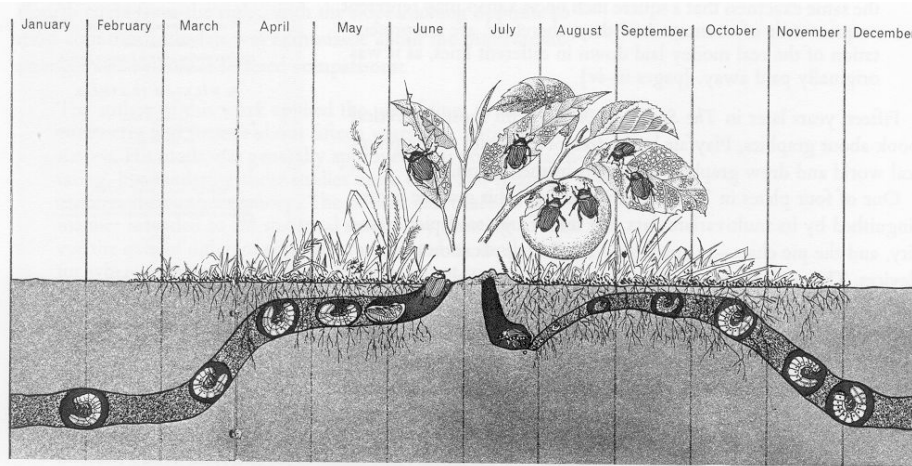
Space-time story

Los Angeles Times, July 22, 1979; based on work of G. McRae, California Institute of technology. (Tufte, 1983, p.42)

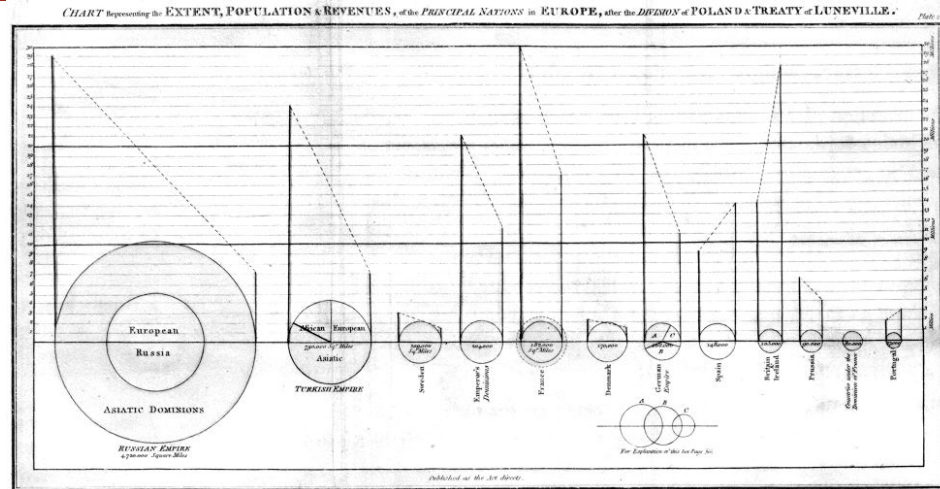


Space-time story

L. H. Newman, Man and Insects (London 1965), p.104-105. (Tufte, 1983, p.43)



Relational Graphics

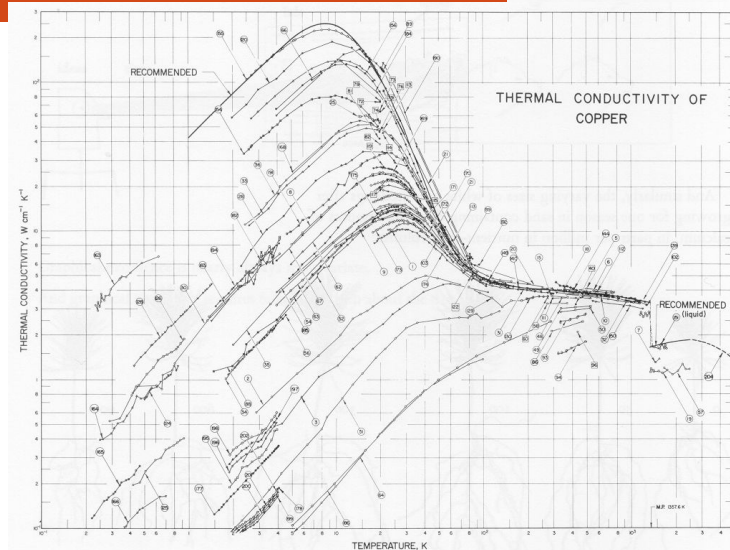


Circles represent area of country, left line: population in millions, right line: taxes in millions (vertical scale). Positive or negative slope is significant but not slope magnitude is not.

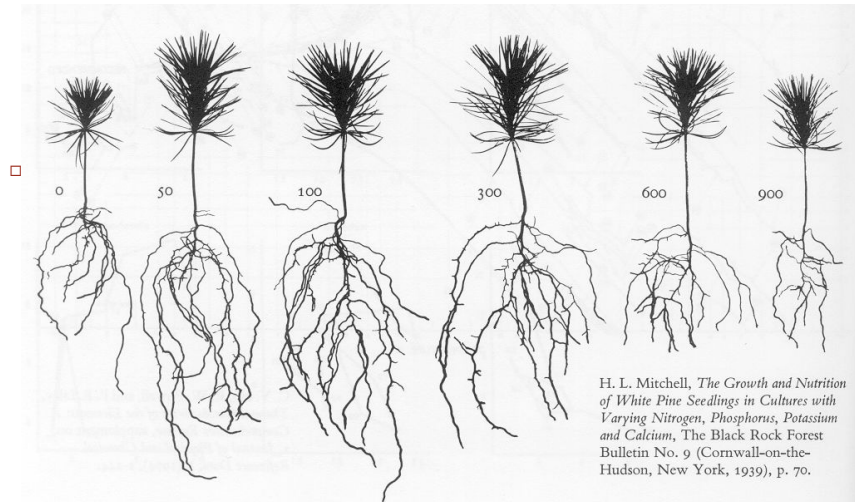
Relational Graphics

C. Y. Ho, R. W. Powell, and P. E. Liley.
Thermal Conductivity of the Elements: A
Comprehensive Review, supplement no.
1, Journal of Physical and Chemical
Reference Data, 3 (1974), 1-244.

- Relationship between temperature and thermal conductivity of copper
- Gathers data from several laboratories
- Makes a clearer and stronger point by the collection
- Connected points are from one publication
- Different answers result from different impurities levels



Relational Graphics



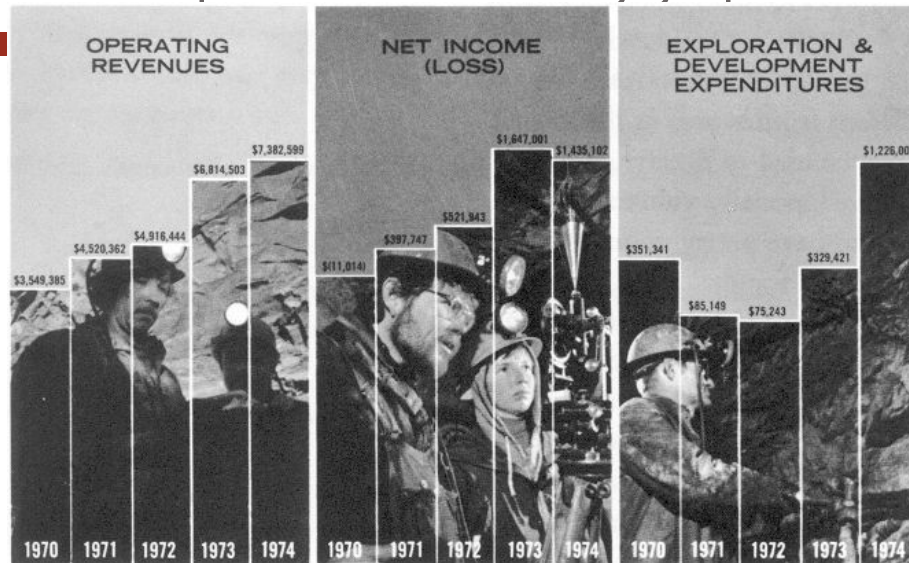
Graphical Excellence - Summary

- Designed for the presentation of interesting data – matter of substance, of statistics, and of design.
- Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
 - > ideas
 - < time
 - < ink
 - < space
- Graphical excellence is nearly always multivariate.
- And is all about truth and integrity.

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'

Examples of misleading graphics



Where is the bottom line? What is happening in 1970?

Day Mines, Inc. 1974 Annual Report, p1 (Tufte, 1983, p54)

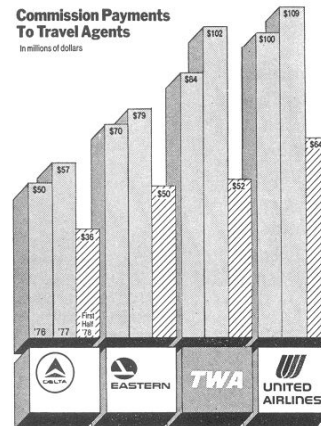
Misleading graphics

New York Times, August 8, 1978, p.D-1 (Tufte, 1983, p54)

What is the first impression of the airlines relative success in 1978?

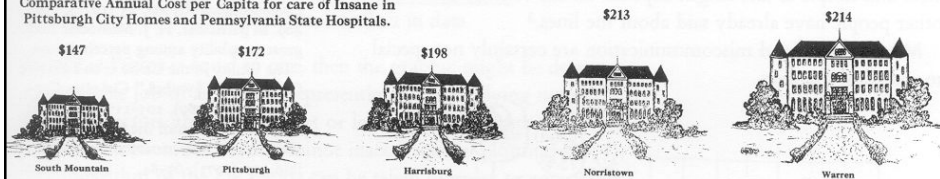
Order of numbers?

Magnitude of numbers? Impression?



Pittsburgh Civic Commission, Report on Expenditures of the Department of Charities (Pittsburgh, 1911), p.7 (Tufte, 1983, p54)

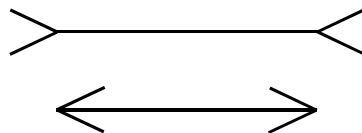
Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.



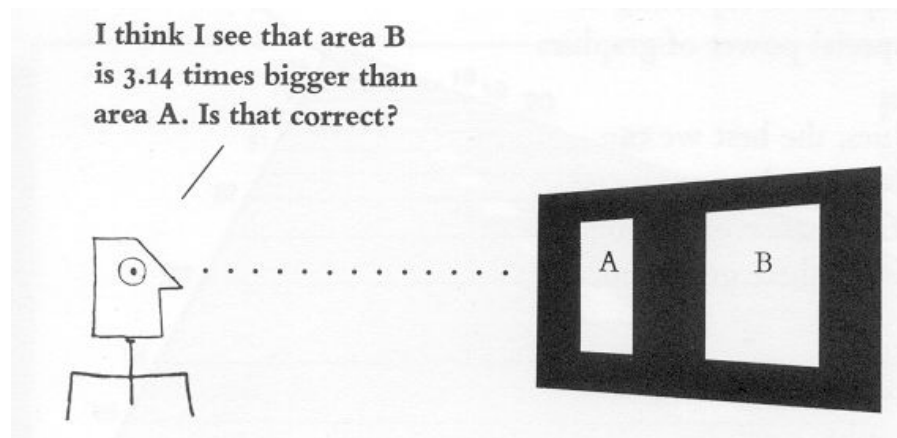
Achieving graphical integrity

- A graphic does not distort if the visual representation is consistent with the numerical representation.
 - Is the magnitude of 'visual representations' as physically measured on the graphic?
 - Or the perceived magnitude?
- Conduct a study of visual perception of the graphics.
 - Circles – perceived area grows more slowly than measured area - reported perceived area = (actual area)*X, where $x = 0.8 \pm 0.3$
- Lines -

□



Achieving graphical integrity



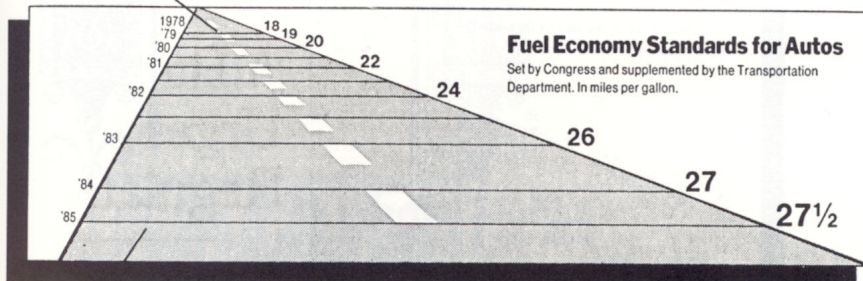
Lie Factors

- Graphic physical measurements = equal data quantities
- Label
- Lie Factor = $\frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$
- Lie factor of 1 – is desirable
- lie factors > 1.05 or < 0.95 go beyond plotting errors

Extreme example

Fuel economy standards for automobiles
 18 miles/gallon in 1978 to 27.5 miles/gallon in 1985
 Increase of 53% = $(27.5 - 18.0)/(18.0) \times 100$

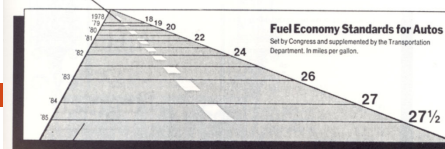
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Extreme example

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

- Graphic increase
- $783\% = (5.3 - 0.6)/(0.6) \times 100$
- Lie Factor = $783/53 = 14.8$
- Additional confounding factors
 - Usually the future is in front of us
 - Dates remain same size and fuel factors increase
 - Includes perspective distortion – how to read change in perspective

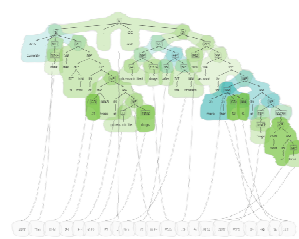
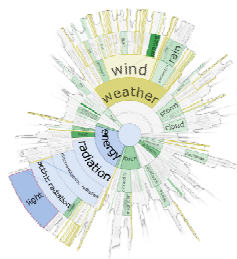
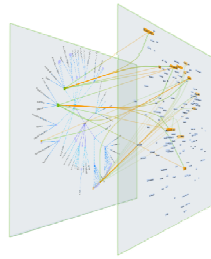
Data Set 2

(((S (S-TPC-2 (NP-SBJ (NP The asbestos fiber)
,)
(VP is (NP crocidolite)
(ADJP-PRD unusually resilient)
(SBAR-TMP once
(S (NP-SBJ it)
(VP enters (NP the lungs))))
{PP with
(S-NOM (NP-SBJ (NP even
(VP causing (PP to brief exposures)
(NP it)))
(NP (NP symptoms)
(SBAR (WHNP-1 that)
(S (NP-SBJ *T*-1)
(VP show (PRT up)
(ADVP-TMP (NP decades
later))))))))))
{NP-SBJ researchers)
(VP said (SBAR o
(S *T*-2)))
(((S (NP-SBJ (NP Lorillard Inc.)
{NP (NP the unit) (PP of (ADJP New York-based)
(NP (NP (NP
(SBAR (WHNP-2 that)
(S (NP-SBJ *T*-2)
(VP makes (NP Kent cigarettes))))))
(VP stopped
(VP using (NP crocidolite)
(PP-LOC-CLR in (NP its Micronite cigarette filters))
(PP-TMP in (NP 1956))))
,))

Data Set 2

(((S (S-TPC-2 (NP-SBJ (NP The asbestos fiber)
,)
(VP is (NP crocidolite)
(ADJP-PRD unusually resilient)
(SBAR-TMP once
(S (NP-SBJ it)
(VP enters (NP the lungs))))
{PP with
(S-NOM (NP-SBJ (NP even
(VP causing (PP to brief exposures)
(NP it)))
(NP (NP symptoms)
(SBAR (WHNP-1 that)
(S (NP-SBJ *T*-1)
(VP show (PRT up)
(ADVP-TMP (NP decades
later))))))))))
{NP-SBJ researchers)
(VP said (SBAR o
(S *T*-2)))
,))

<http://www.cs.toronto.edu/~gpenn/esslii09>



CHRISTOPHER COLLINS
CCOLLINS@CS.UTORONTO.CA

GERALD PENN
GPENN@CS.UTORONTO.CA

SHEELAGH CARPENDALE
SHEELAGH@UCALGARY.CA

