Visualising the difference: revealing pattern and structure through graphical techniques

Tony Hirst
Dept of Communication and Systems
The Open University

Metalogue: Why Do Things Get in a Muddle?*

Daughter: Daddy, why do things get in a muddle?

Father: What do you mean? Things? Muddle?

D: Well, people spend a lot of time tidying things, but they never seem to spend time muddling them. Things just seem to get in a muddle by themselves. And then people have to tidy them up again.

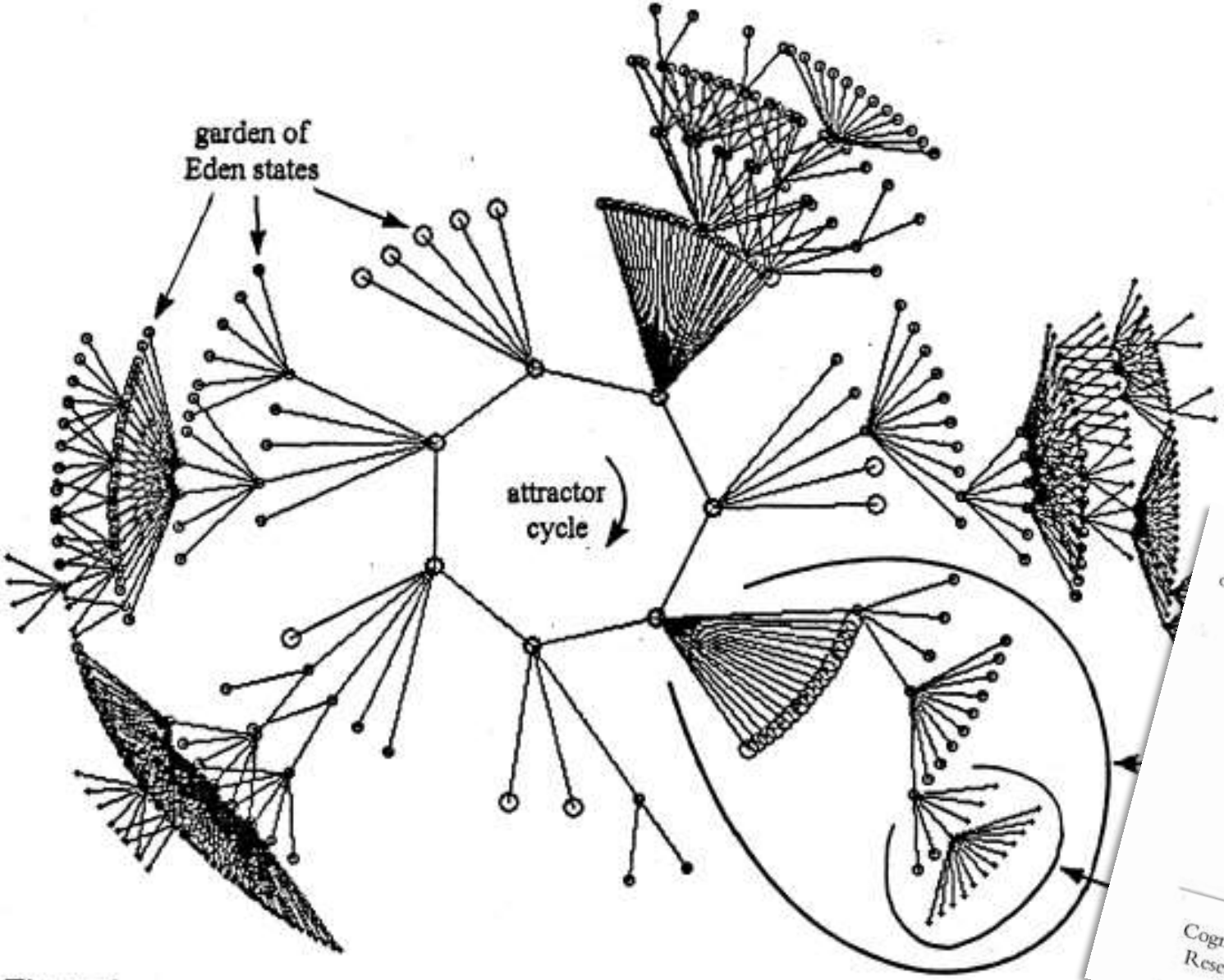
F: But do your things get in a muddle if you don't touch them?

D: No—not if *nobody* touches them. But if you touch them—or if anybody touches them—they get in a muddle and it's a worse muddle if it isn't me.

F: Yes—that's why I try to keep you from touching the things on my desk. Because my things get in a worse muddle if they are touched by somebody who isn't *me*.

D: But do people *always* muddle other people's things? Why do they, Daddy?

F: Now, wait a minute. It's not so simple. First of all, what do you mean by a muddle?




The Ghost in the Machine
Basins of Attraction
of Random Boolean Networks

Andrew Wuensche

CSRP 281

1993

UNIVERSITY OF
SUSSEX
AT BRIGHTON

Cognitive Science
Research Papers

Figure 1

Above. *A basin of attraction of a random Boolean network (N=13, K=3). The basin links 604 states, of which 523 are garden of Eden states. The attractor has period 7. The direction of time is inwards from garden of Eden states to the attractor, then clock-wise. The basin is one of 15, and is indicated in the basin of attraction field in*

cell	*wiring*	*rule,*	*table*
1	3,12,6	86,	01010110
2	7,11,4	4,	00000100
3	3,3,1	196,	11000100
4	11,3,9	52,	00110100

Visual Analysis

vs.

Presentation Graphics

Hans Rosling: No more boring data: TEDTalks

TEDtalksDirector 981 videos Subscribe

Like Add to Share 537,082

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

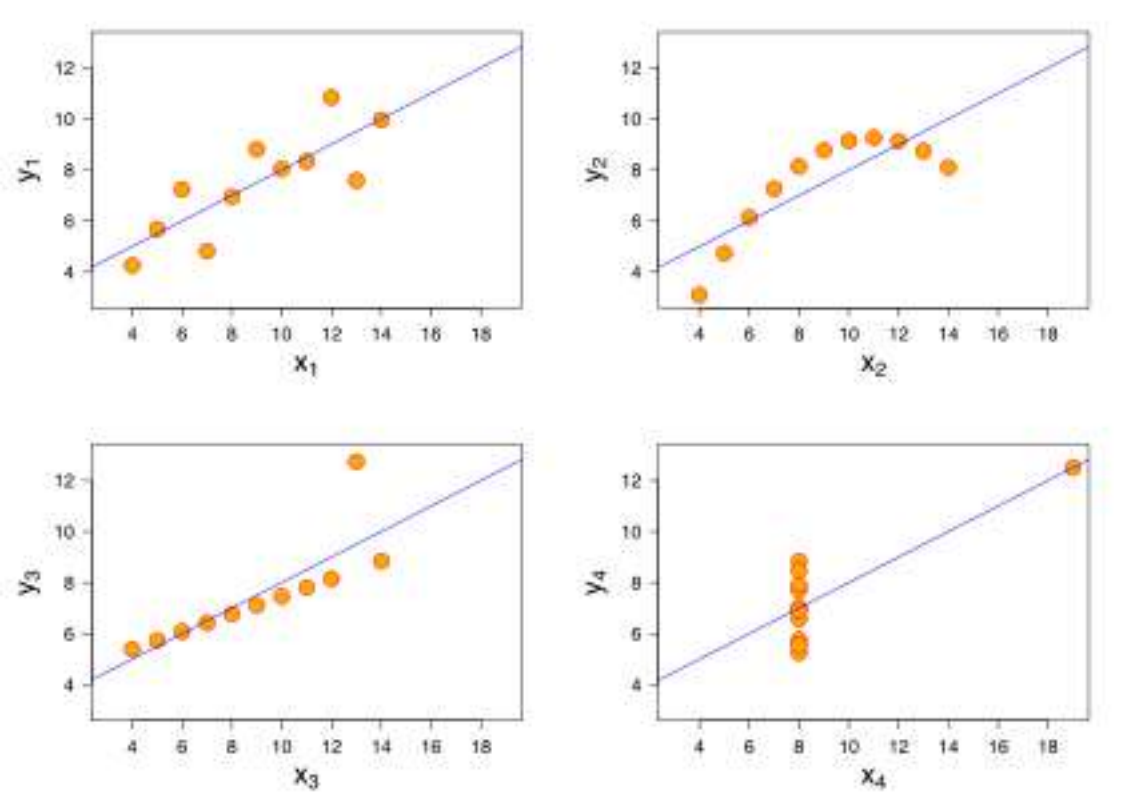
Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.1				
13.0	7.58	13.0	8.7				
9.0	8.81	9.0	8.7				
11.0	8.33	11.0	9.2				
14.0	9.96	14.0	8.1				
6.0	7.24	6.0	6.1				
4.0	4.26	4.0	3.1				
12.0	10.84	12.0	9.1				
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Property	Value
Mean of x in each case	9 exact
Variance of x in each case	10 exact
Mean of y in each case	7.50 (to 2 d.p.)
Variance of y in each case	3.75 (to 2 d.p.)
Correlation between x and y in each case	0.816 (to 3 d.p.)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 d.p. and 3 d.p. resp.)

Anscombe's Quartet

I		II		III		
x	y	x	y	x	y	x
10.0	8.04	10.0	9.14	10.0	7.46	8.
8.0	6.95	8.0	8.14	8.0	6.77	8.
13.0	7.58	13.0	8.74	13.0	12.74	8.
9.0	8.81	9.0	8.77	9.0	7.11	8.
11.0	8.33	11.0	9.26	11.0	7.81	8.
14.0	9.96	14.0	8.10	14.0	8.84	8.
6.0	7.24	6.0	6.13	6.0	6.08	8.
4.0	4.26	4.0	3.10	4.0	5.39	19
12.0	10.84	12.0	9.13	12.0	8.15	8.
7.0	4.82	7.0	7.26	7.0	6.42	8.
5.0	5.68	5.0	4.74	5.0	5.73	8.



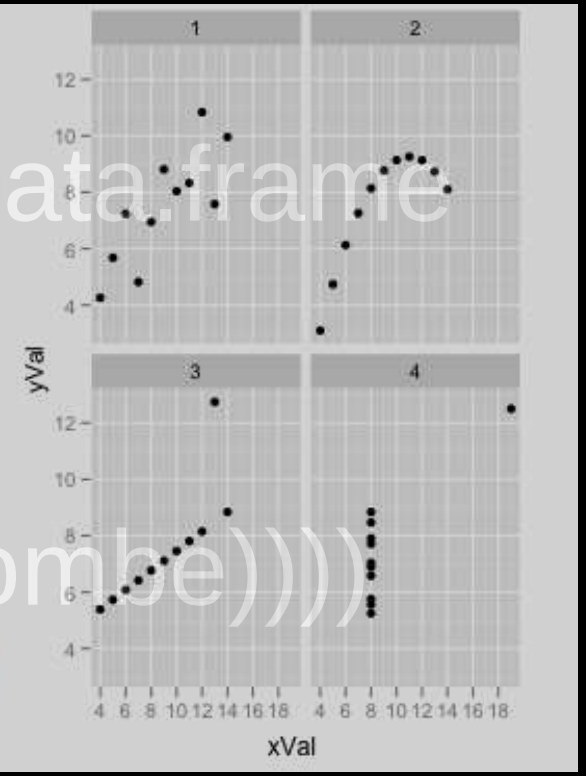
```
library(ggplot2)

mydata=with(anscombe,data.frame
(xVal=c(x1,x2,x3,x4),
yVal=c(y1,y2,y3,y4),
mygroup=gl(4,nrow(anscombe))))

ggplot(mydata,aes(x=xVal,y=yVal))
+geom_point()+facet_wrap(~mygro
up)
```

library(ggplot2)

```
> library("ggplot2")
> mydata=with(anscombe,data.frame(xVal=c(x1,x2,x3,x4), yVal=c(y1,y2,y3,y4), mygroup=gl(
> ggplot(mydata,aes(x=xVal,y=yVal))+geom_point()+facet_wrap(~mygroup)
```



mydata=with(anscombe,data.frame(xVal=c(x1,x2,x3,x4), yVal=c(y1,y2,y3,y4), mygroup=gl(4,nrow(anscombe))))

ggplot(mydata,aes(x=xVal,y=yVal))+geom_point()+facet_wrap(~mygroup)

Information required to generate a visualisation

VS

Information revealed by a visualisation

URL: (*Link to this page*)

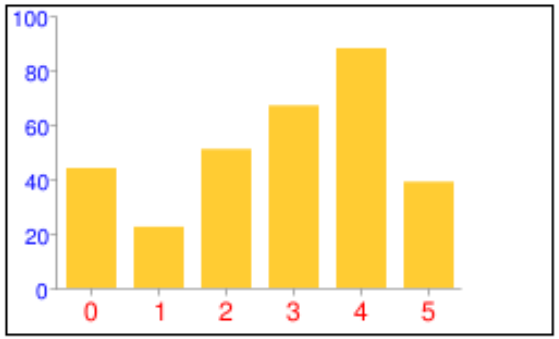
```
http://chart.googleapis.com/chart?cht=bvg&chs=250x150&chd=t:44,22.45,51,67,88,39&chxt=x,y&chxs=0,ff0000,12,0,lt|1,0000ff1
```

Parameters:

```
cht=bvg
chs=250x150
chd=t:44,22.45,51,67,88,39
chxt=x,y
chxs=0,ff0000,12,0,lt
     1,0000ff,10,1,lt
```

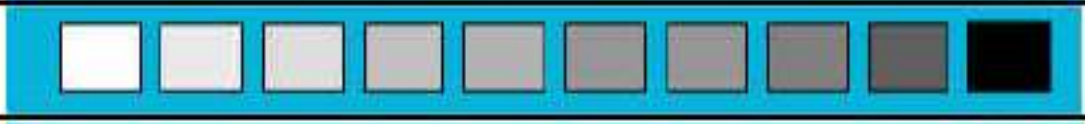
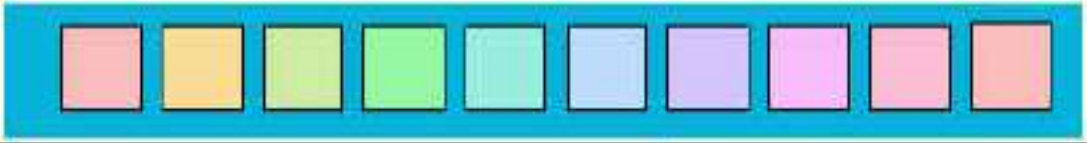
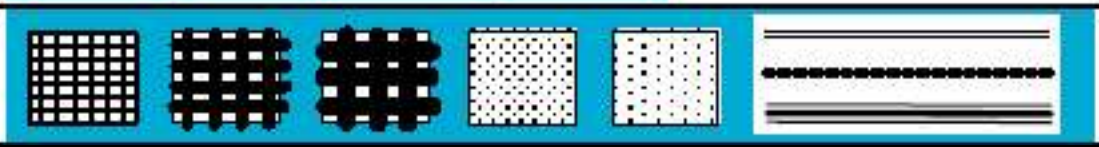
Reformat Parameters

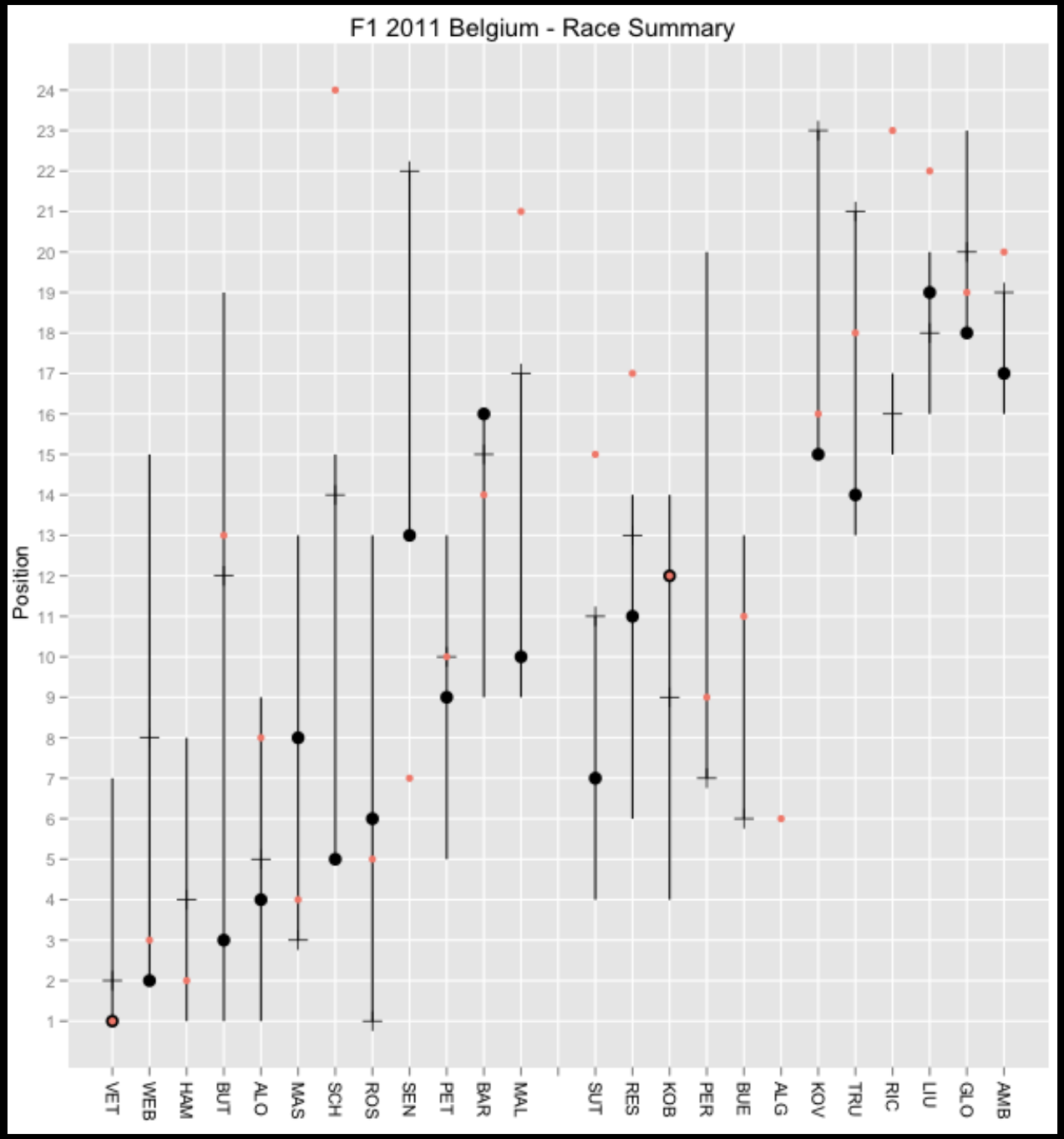
Generated Chart:



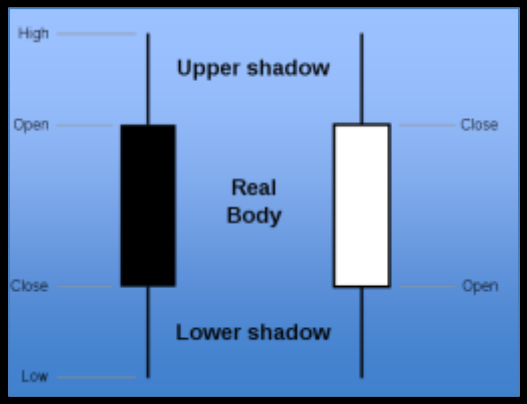
Show errors...

Bertin's Original Visual Variables

Position changes in the x, y location	
Size change in length, area or repetition	
Shape infinite number of shapes	
Value changes from light to dark	
Colour changes in hue at a given value	
Orientation changes in alignment	
Texture variation in 'grain'	


F1 2011 Belgium - Race Summary
Position
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
VET
WEB
HAM
BUT
ALO
MAS
SCH
ROS
SEN
PET
BAR
MAL
SUT
RES
KOB
PER
BUE
ALG
KOV
TRU
RIC
LIU
GLO
AMB


1.2630
1.2620
1.2610
1.2600
1.2590
1.2580
1.2570
1.2560
1.2550
05:00
07:30
10:00
12:15
14:30
16:45


High
Upper shadow
Open
Close
Real
Body
Close
Open
Lower shadow
Low

Visualisations can make structure evident

Google docs MP expenses, 07-08 Anyone with the link View only Share

File Edit View Insert Format Data Tools Help

Formula: Show all formulae

	B	C	D	E	F	G	H	I	J	K	L
1	BACKGROUNI COLOUR CODE	PINK = CLAIMING FULL ADDITIONAL COSTS ALLOWANCE	BLUE = CLAIMING NO ADDITIONAL COSTS ALLOWANCE, EVEN THOUGH ENTITLED					DATALINK: ALL THE LATEST REVELATIONS	http://spreadsheets key=rvWgEEGK9x		
2	Member		Party	Constituency	Maj	% swing needed to unseat	LINK TO EXPENSES DETAILS	TOTAL ALLOWANCES CLAIMED, INC TRAVEL	TOTAL BASIC ALLOWANCES, EXC TRAVEL	TOTAL TRAVEL CLAIMED	ADDITIONAL COSTS ALLOWANCE (Cost of staying away from main
3	Abbott	Ms Diane	Labour	Hackney North & Stoke Newington				131735			0
4	Adams	Mr Gerry	Sinn Fein	West Belfast	19315						21131
5	Afriyie	Adam	Conservative	Windsor							0
6	Ainger	Nick	Labour	Carmarthen West & Pembrokeshire South	1910						95
7	Ainsworth	Mr Peter	Conservative	Surrey East	15921						3
8	Ainsworth	Rt Hon Bob	Labour	Coventry North East	14222						
9				Inverness, Nairn							

DATABLOG
Facts are sacred

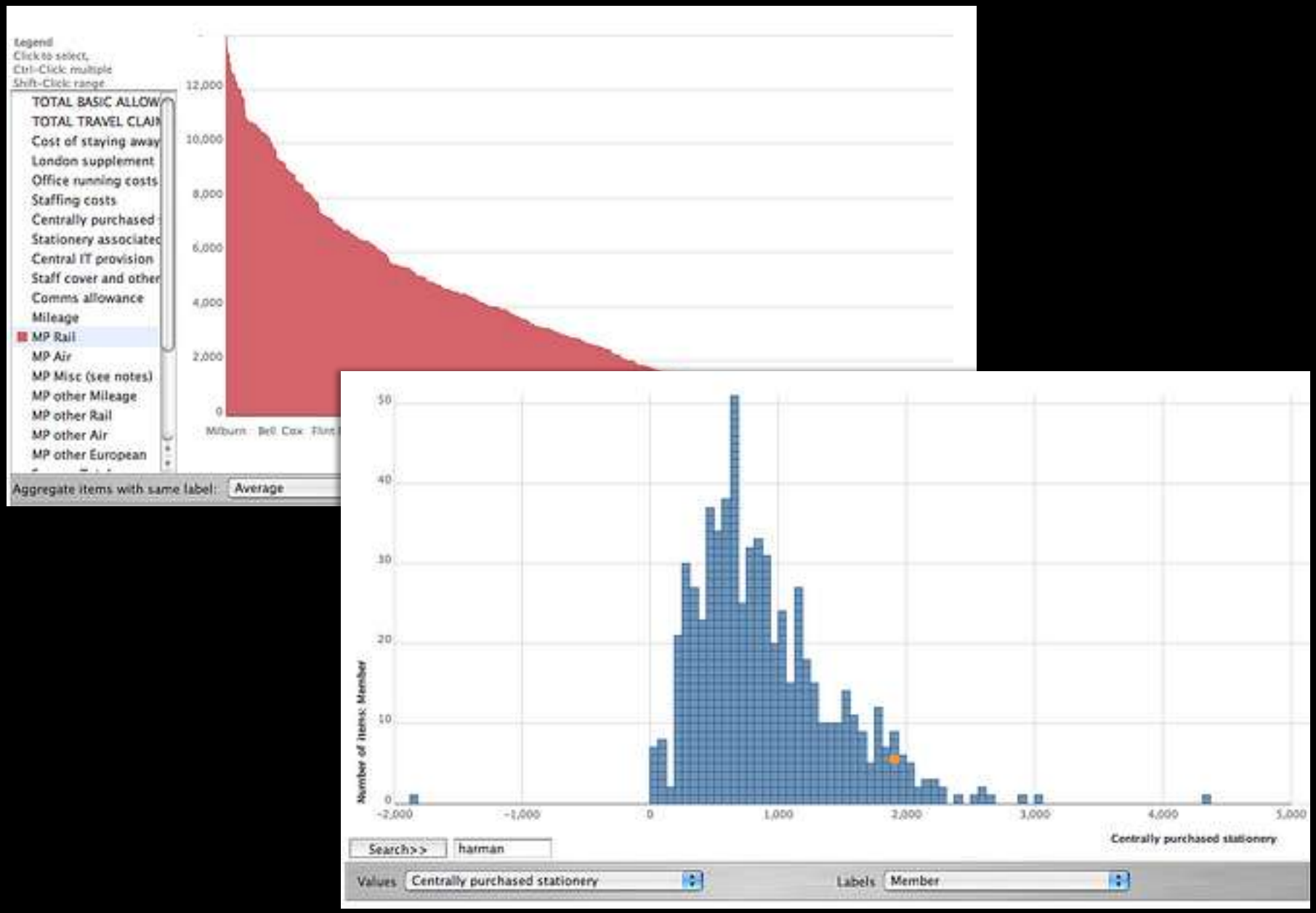
Blog home

Previous

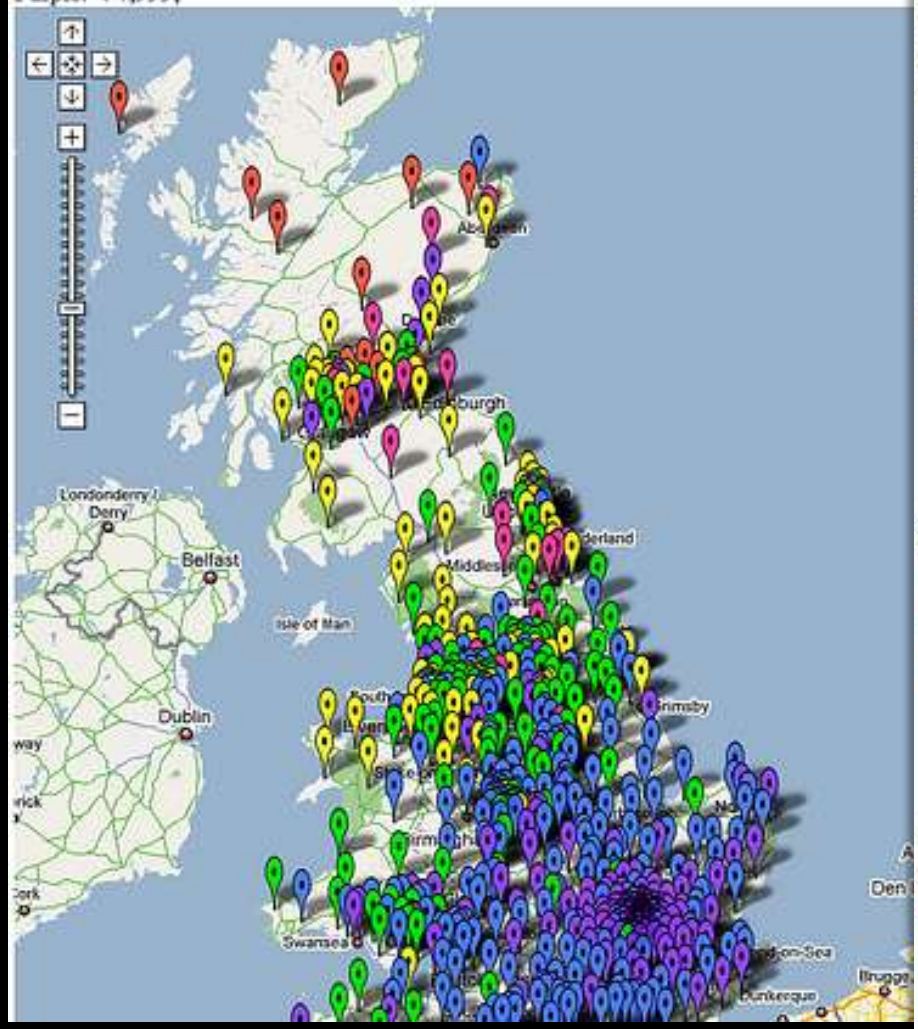
MPs' expenses: who claimed what?

We bring you the full list - in a form you can use

- Get the 2008/09 figures


Legend
Click to select,
Ctrl-Click multiple
Shift-Click range
TOTAL BASIC ALLOW
TOTAL TRAVEL CLAI
Cost of staying away
London supplement
Office running costs
Staffing costs
Centrally purchased
Stationery associated
Central IT provision
Staff cover and other
Comms allowance
Mileage
MP Rail
MP Air
MP Misc (see notes)
MP other Mileage
MP other Rail
MP other Air
MP other European
12,000
10,000
8,000
6,000
4,000
2,000
0
Milburn
Bell
Cox
Aggregate items with same label:
Average
Number of items: Member
50
40
30
20
10
0
-2,000
-1,000
0
1,000
2,000
3,000
4,000
5,000
Centrally purchased stationery
Search>>
harman
Values
Centrally purchased stationery
Labels
Member

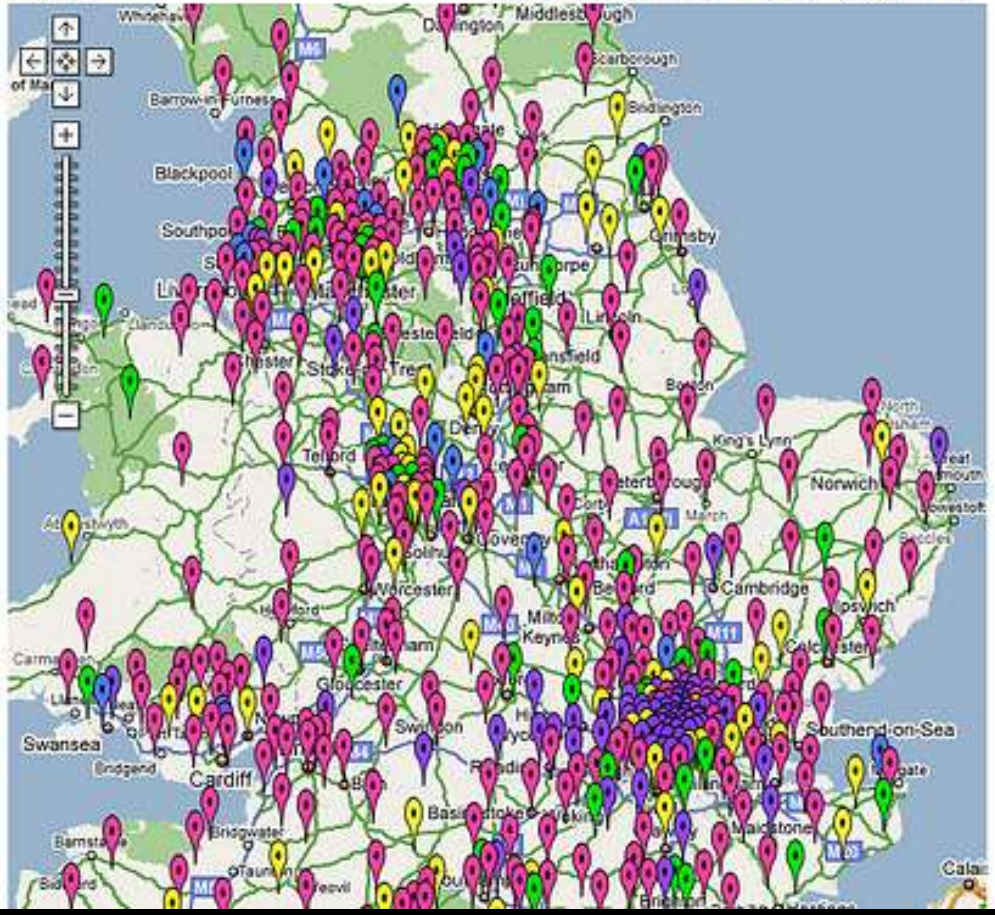
Map markers show MP Total Claimed Travel Exxpenses over 2007-8 in different bands:
: Red: > 25,000
Pink: 20,000-24,999;
Yellow: 15,000-19,999;
Green: 10,000-14,999;
Blue: 5,000-9,999;
Purple: < 4,999;



clear | Total Cost of Staying Away from Main Home

Marker values:

Travel Type	Red	Pink	Yellow	Green	Blue	Purple
Total Travel Claimed	>25k	20k-25k	15k-20k	10k-15k	5k-10k	0-5k
Total Costs of Living Away From Main Home	>25k	20k-25k	15k-20k	10k-15k	5k-10k	0-5k
Air	>5k	4k-5k	3k-4k	2k-3k	1k-2k	0-1k
Rail	>10k	8k-10k	6k-8k	4k-6k	2k-4k	0-2k
Mileage	>10k	8k-10k	6k-8k	4k-6k	2k-4k	0-2k



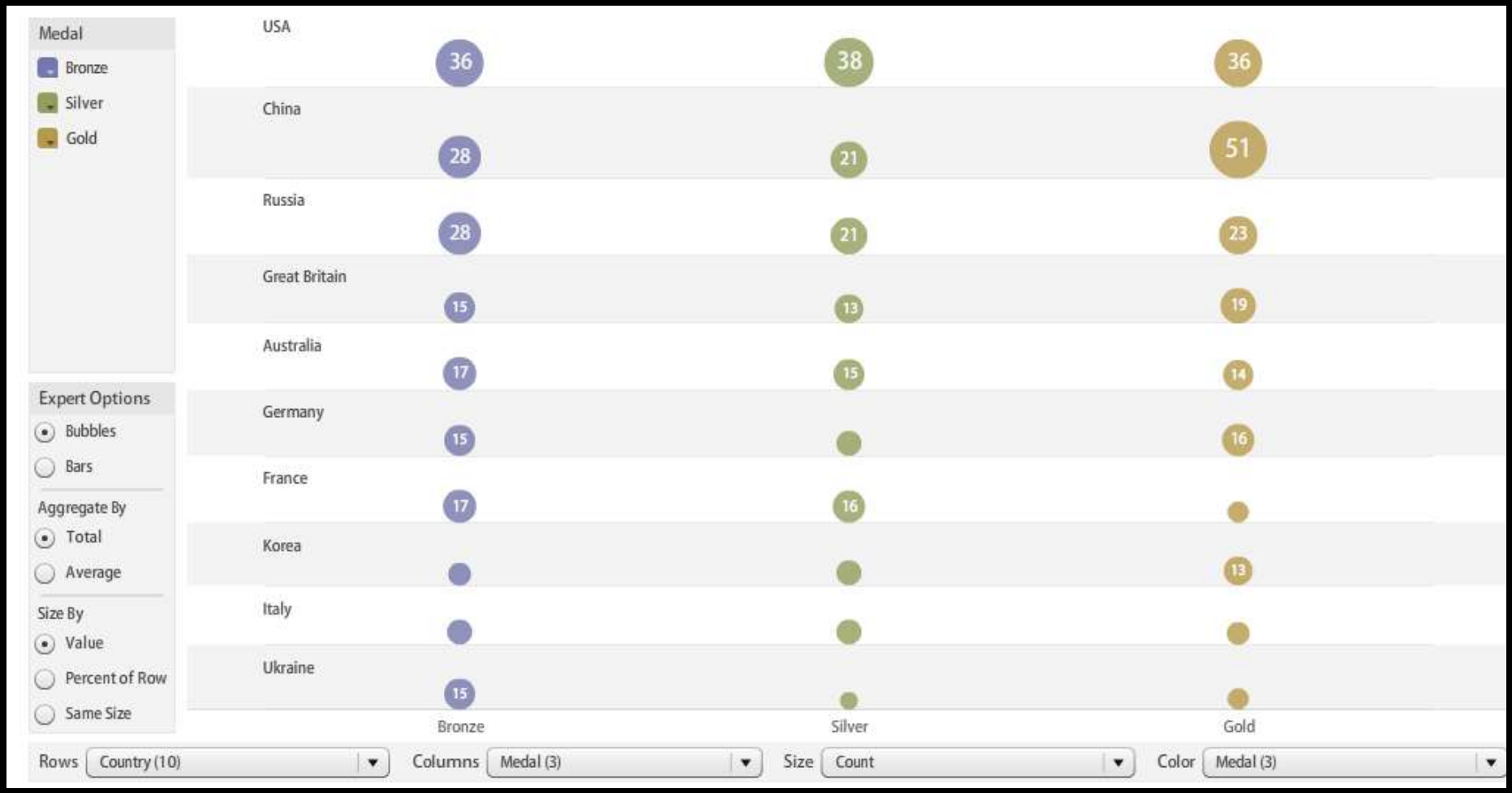
Variable encoding:
Data variable -> graphical dimension

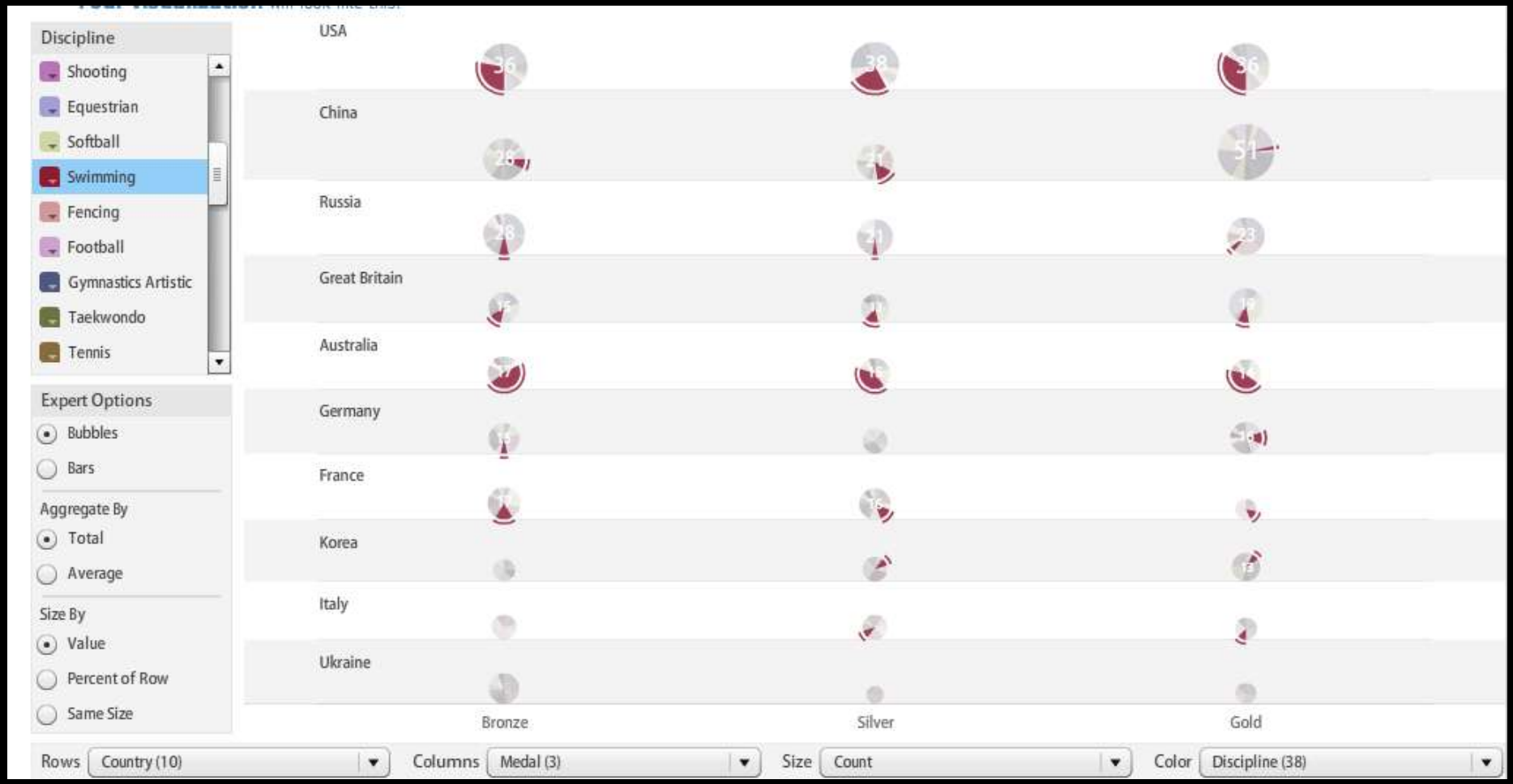
BUT…

To what extent does the viewer use the visualisation to inform the creation of a model that they then interpret in order to spot the differences that make a difference in the visualisation?

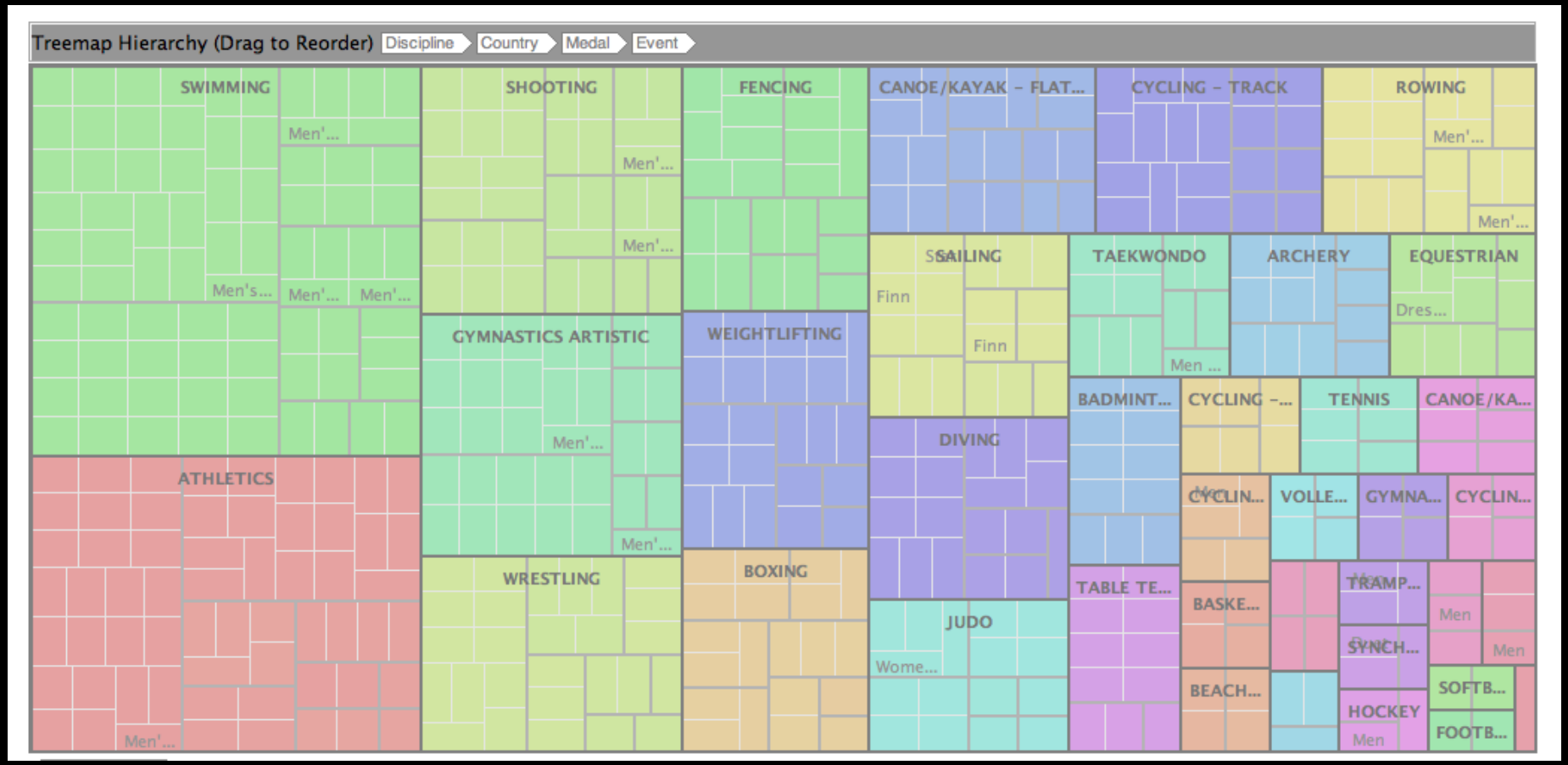
Seeing Structure in Tabular Data

	Country	Discipline	Event	Medal	Name
1	USA	Athletics	Men's 100m	Bronze	DIX Walter
2	USA	Athletics	Men's 200m	Silver	CRAWFORD Shawn
3	USA	Athletics	Men's 200m	Bronze	DIX Walter
4	USA	Athletics	Men's 400m	Gold	MERRITT LaShawn
5	USA	Athletics	Men's 400m	Silver	WARINER Jeremy
6	USA	Athletics	Men's 400m	Bronze	NEVILLE David
7	USA	Athletics	Men's 110m Hurdles	Silver	PAYNE David
8	USA	Athletics	Men's 110m Hurdles	Bronze	OLIVER David
9	USA	Athletics	Men's 400m Hurdles	Gold	TAYLOR Angelo
10	USA	Athletics	Men's 400m Hurdles	Silver	CLEMENT Kerron
11	USA	Athletics	Men's 400m Hurdles	Bronze	JACKSON Bershawn
12	USA	Athletics	Men's Shot Put	Silver	CANTWELL Christian
13	USA	Athletics	Men's Decathlon	Gold	CLAY Bryan
14	USA	Athletics	Women's 10,000m	Bronze	FLANAGAN Shalane

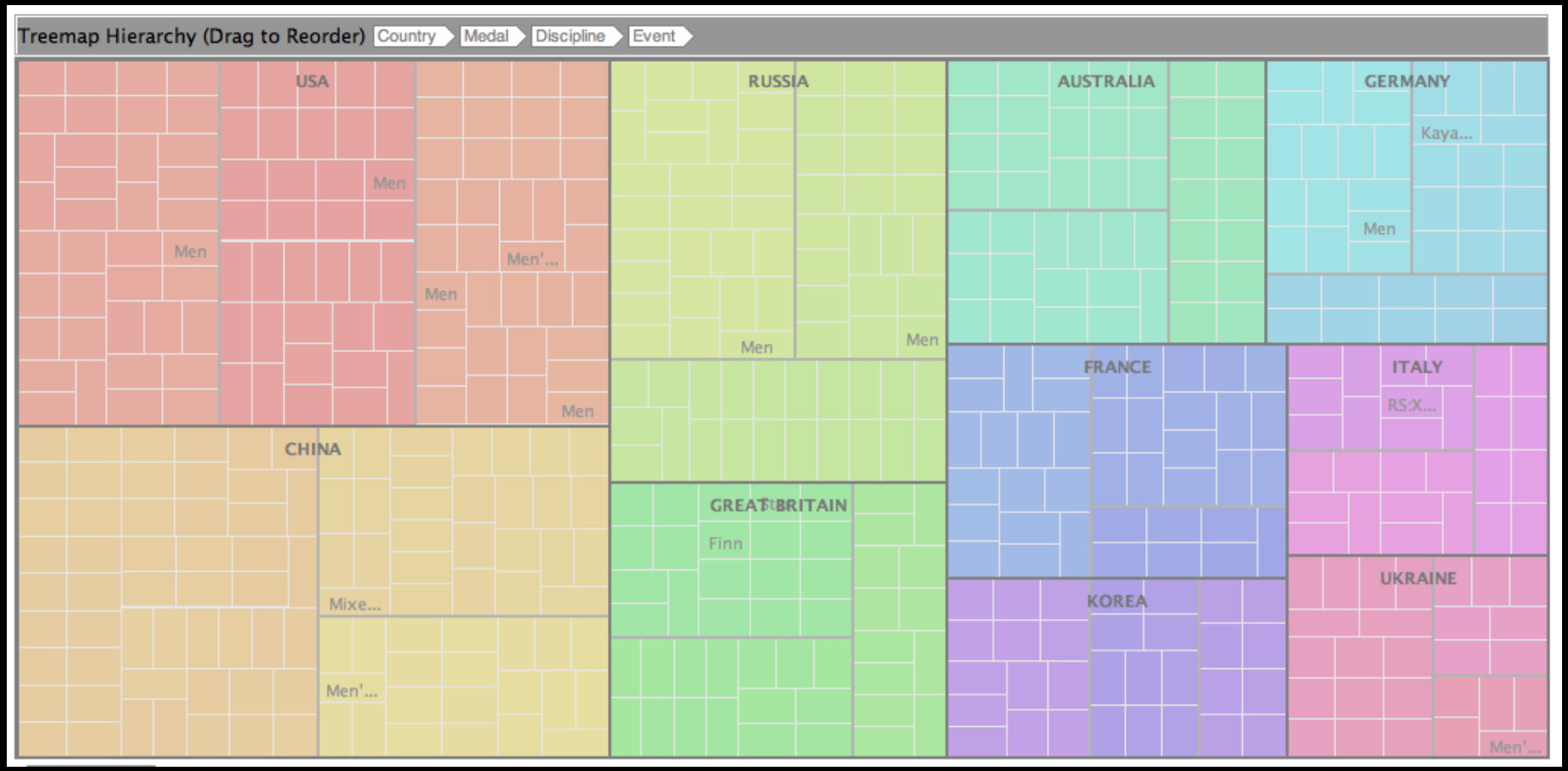

Medal
Bronze
Silver
Gold
Expert Options
Bubbles
Bars
Aggregate By
Total
Average
Size By
Value
Percent of Row
Same Size
USA
36
38
36
China
28
21
51
Russia
28
21
23
Great Britain
15
13
19
Australia
17
15
14
Germany
15
16
France
17
16
Korea
13
Italy
Ukraine
15
Bronze
Silver
Gold
Rows
Country (10)
Columns
Medal (3)
Size
Count
Color
Medal (3)


Discipline
Shooting
Equestrian
Softball
Swimming
Fencing
Football
Gymnastics Artistic
Taekwondo
Tennis
Expert Options
Bubbles
Bars
Aggregate By
Total
Average
Size By
Value
Percent of Row
Same Size
USA
China
Russia
Great Britain
Australia
Germany
France
Korea
Italy
Ukraine
Bronze
Silver
Gold
Rows
Country (10)
Columns
Medal (3)
Size
Count
Color
Discipline (38)

	Country	Discipline	Event	Medal	Name
1	USA	Athletics	Men's 100m	Bronze	DIX Walter
2	USA	Athletics	Men's 200m	Silver	CRAWFORD Shawn
3	USA	Athletics	Men's 200m	Bronze	DIX Walter
4	USA	Athletics	Men's 400m	Gold	MERRITT LaShawn
5	USA	Athletics	Men's 400m	Silver	WARINER Jeremy
6	USA	Athletics	Men's 400m	Bronze	NEVILLE David
7	USA	Athletics	Men's 110m Hurdles	Silver	PAYNE David
8	USA	Athletics	Men's 110m Hurdles	Bronze	OLIVER David
9	USA	Athletics	Men's 400m Hurdles	Gold	TAYLOR Angelo
10	USA	Athletics	Men's 400m Hurdles	Silver	CLEMENT Kerron
11	USA	Athletics	Men's 400m Hurdles	Bronze	JACKSON Bershawn
12	USA	Athletics	Men's Shot Put	Silver	CANTWELL Christian
13	USA	Athletics	Men's Decathlon	Gold	CLAY Bryan
14	USA	Athletics	Women's 10,000m	Bronze	FLANAGAN Shalane



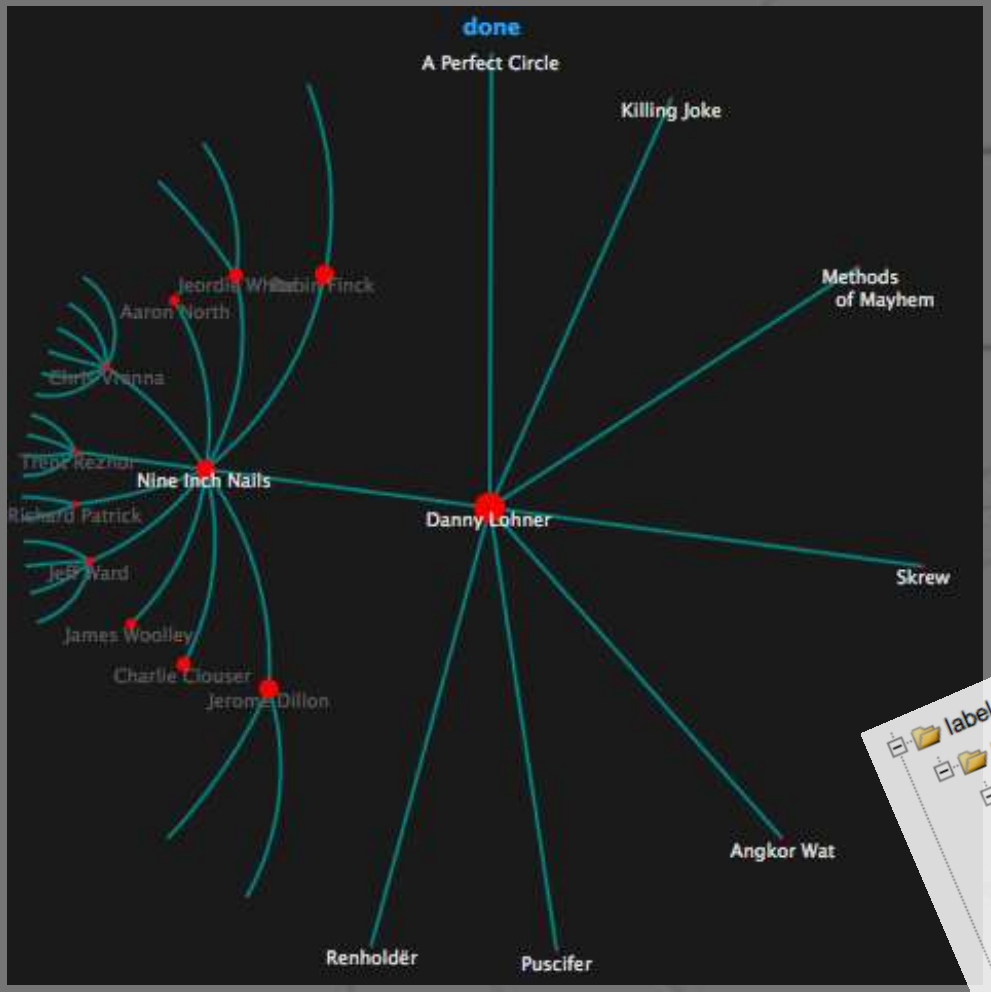
Treemap Hierarchy (Drag to Reorder) Discipline Country Medal Event

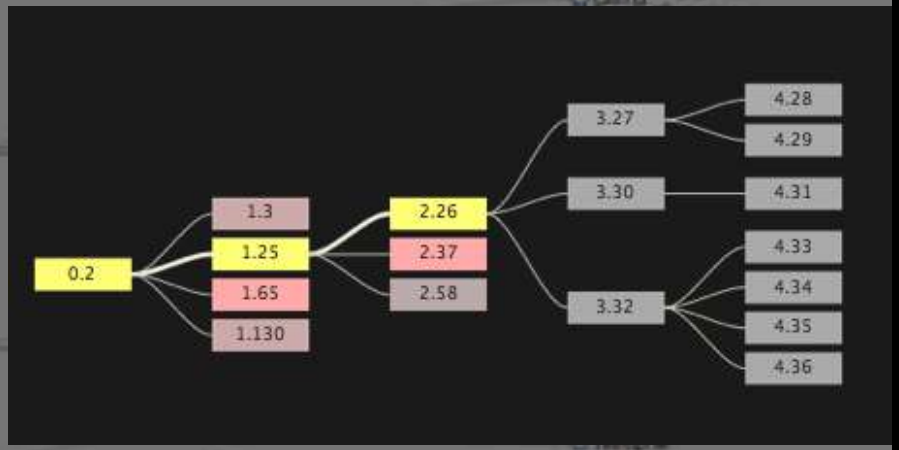


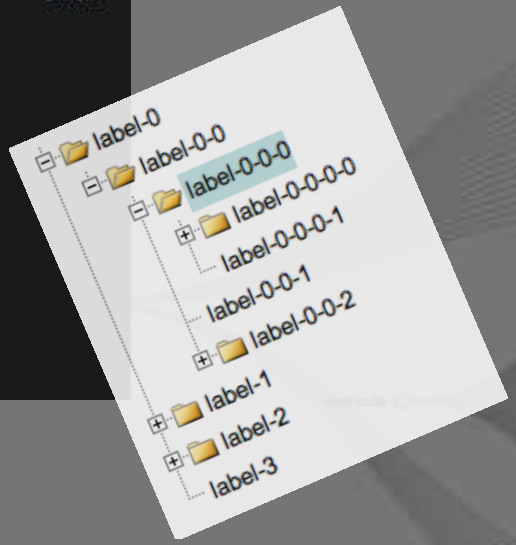
Treemap Hierarchy (Drag to Reorder) Country Medal Discipline Event

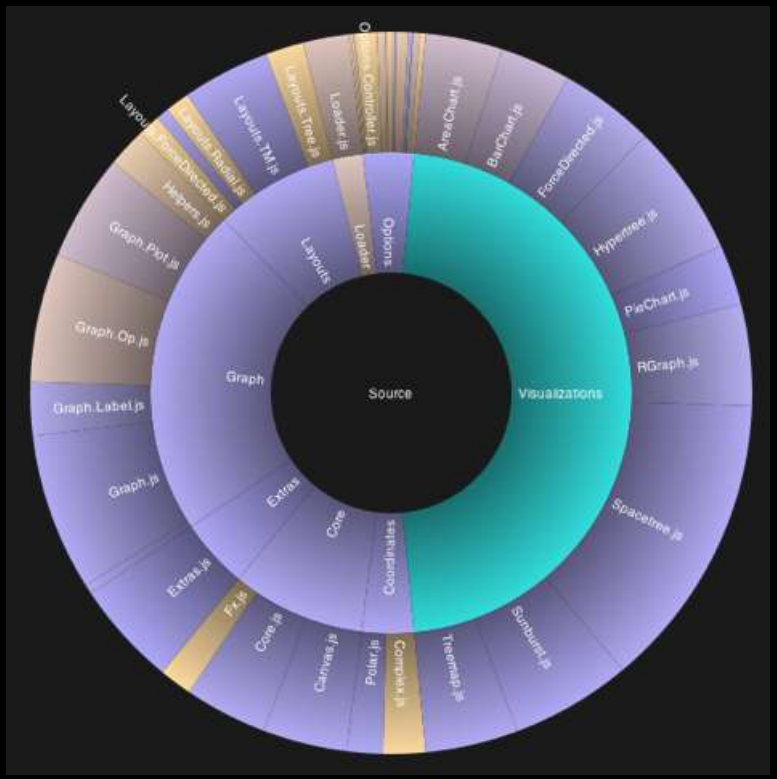
Trees: levels or containers?

d3.js Dendrogram



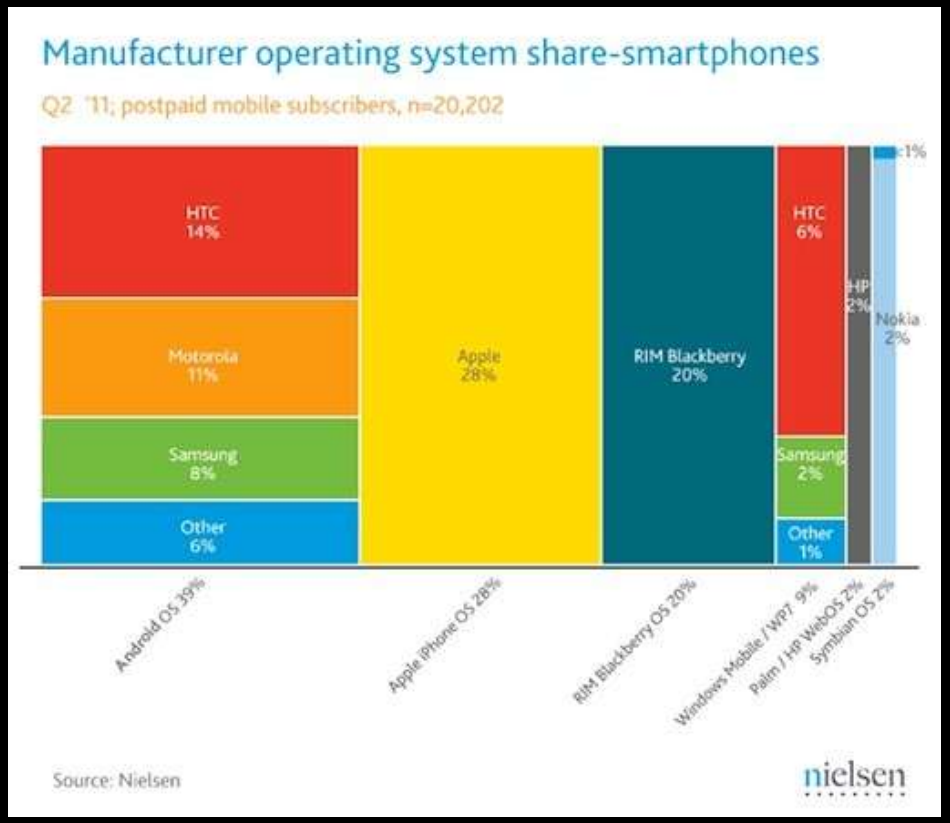


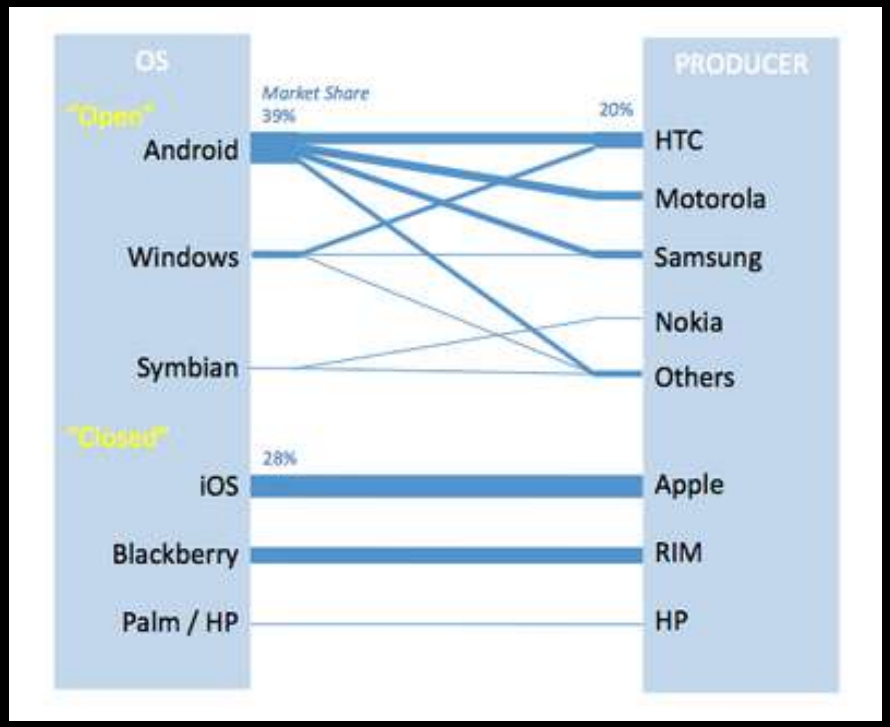




(implied) containment

When you get the structure wrong….
Marimekko/mosaic charts vs flow chart





http://bit.ly/qhZfbB

http://junkcharts.typepad.com/junk_charts/2011/08/false-promises-of-equality-and-structure.html

2

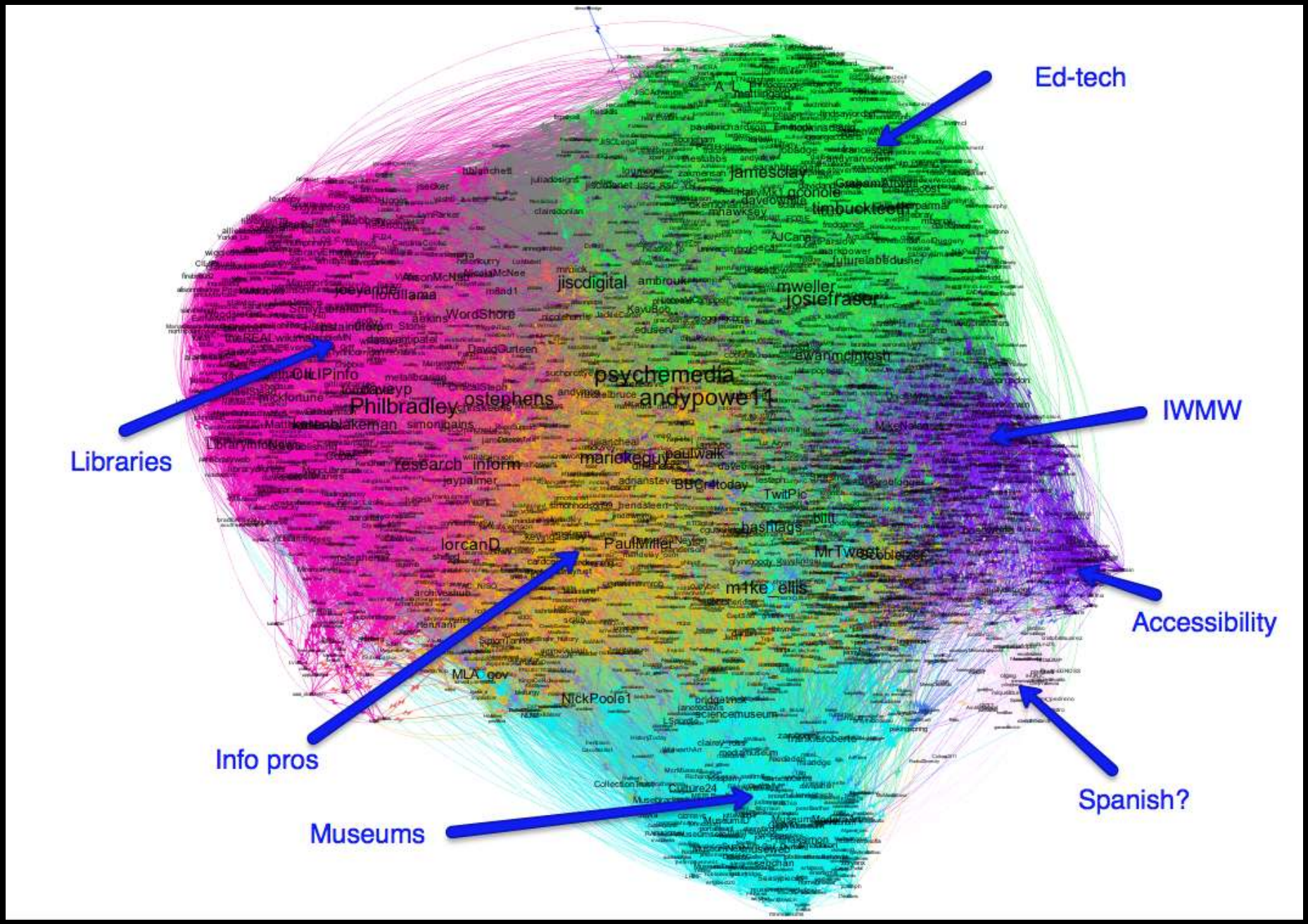
X

7

seven data types (
1-, 2-, 3-dimensional data,
temporal and multi-dimensional data,
tree and network data)

seven tasks (
overview,
zoom,
filter,
details-on-demand,
relate,
history,
extract)

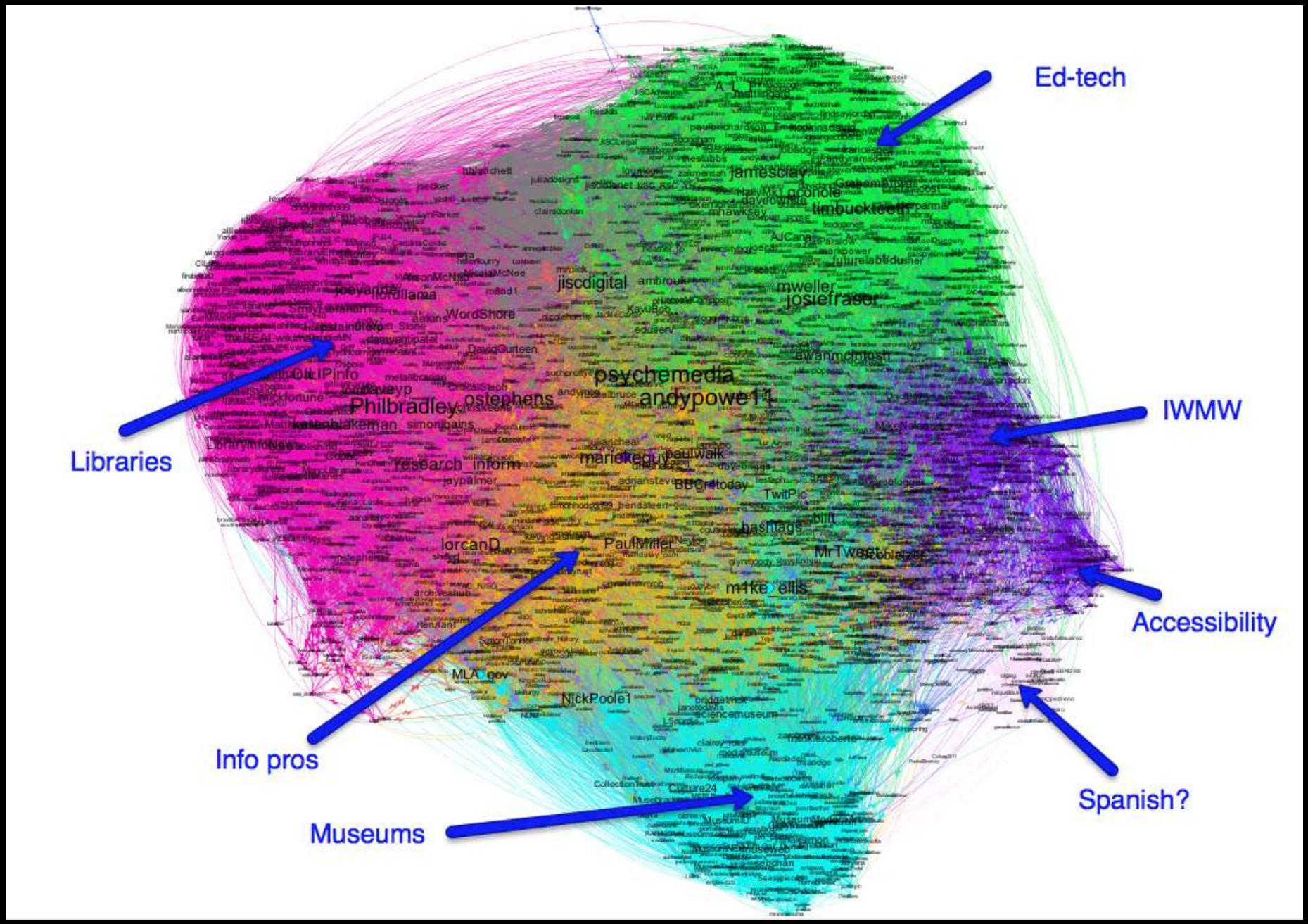
From: *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*
Ben Schneiderman, IEEE Symposium on Visual Languages, 1996

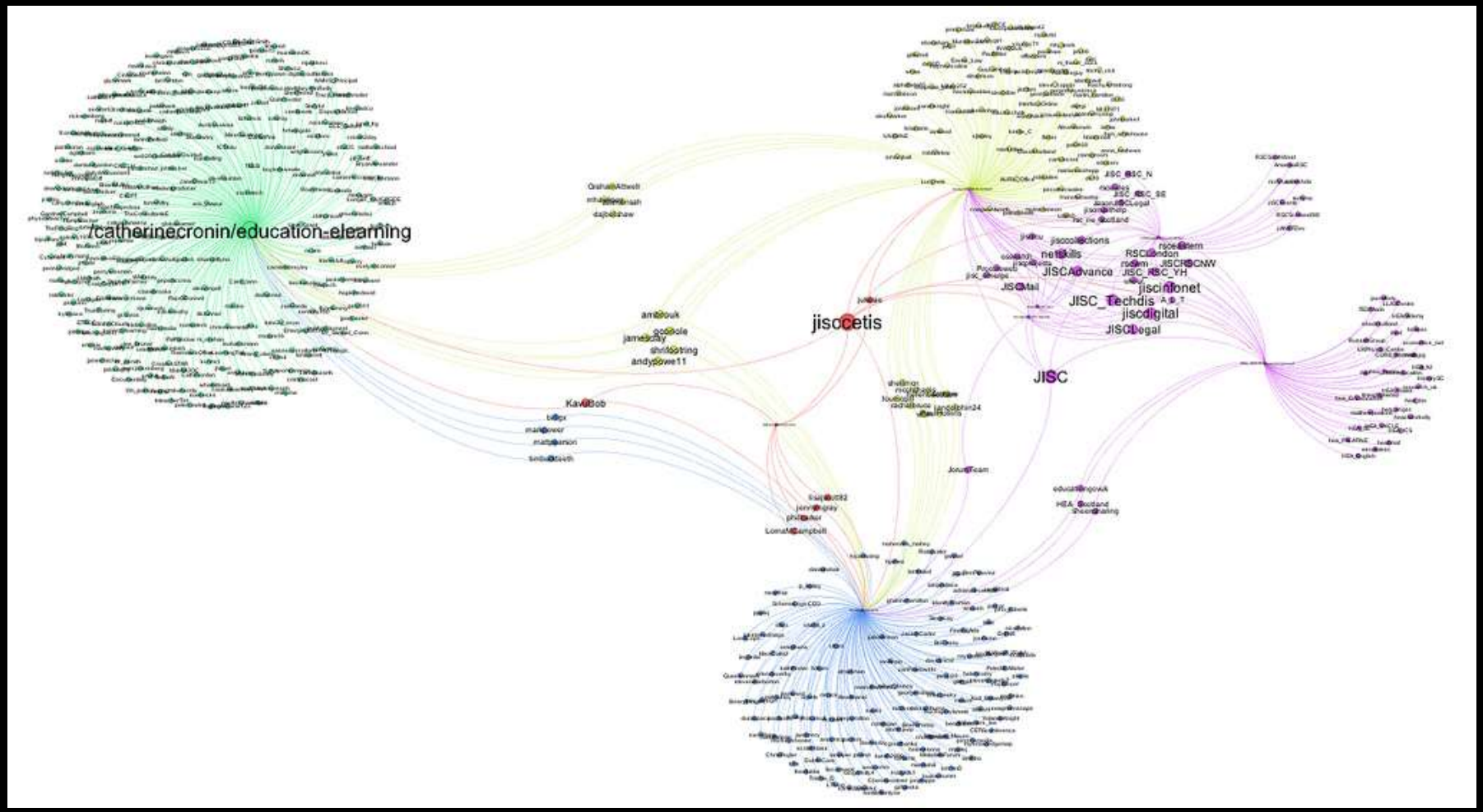

Ed-tech
IWMW
Libraries
Accessibility
Info pros
Spanish?
Museums
jamesclay
timbucktoo
jiscdigital
mweller
josiefraser
WordShore
psychemedia
andypowe11
Philbradley
ostephens
mariekeguy
paulwalk
research
inform
TwitPic
hashtags
lorcanD
PaulMiller
MrTweet
m1ke_ellis
MLA_gov
NickPoole1
Culture24

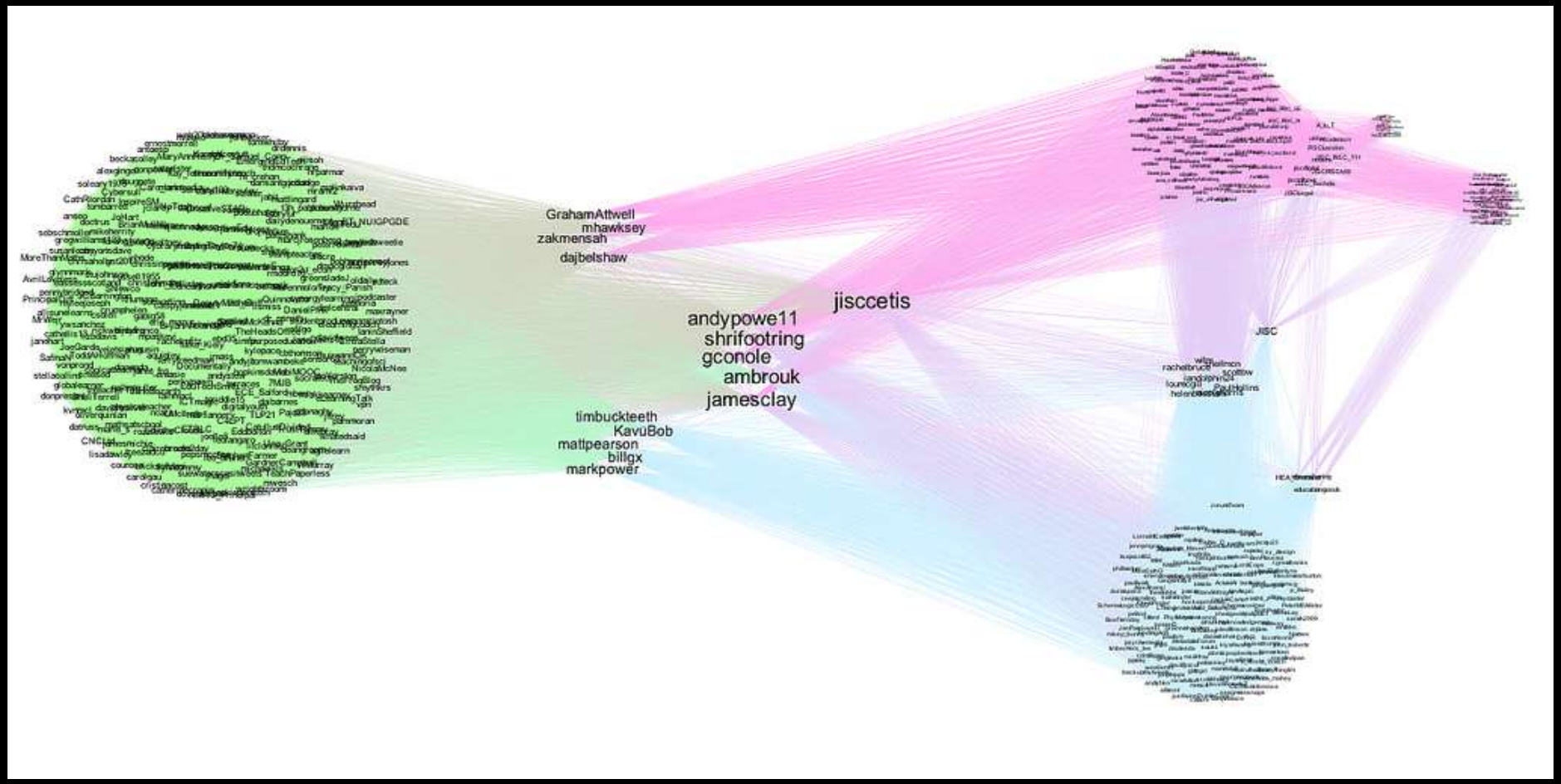
SCHNEIDERMAN'S "VISUAL INFORMATION SEEKING MANTRA"

Overview first, zoom and filter, then details-on-demand

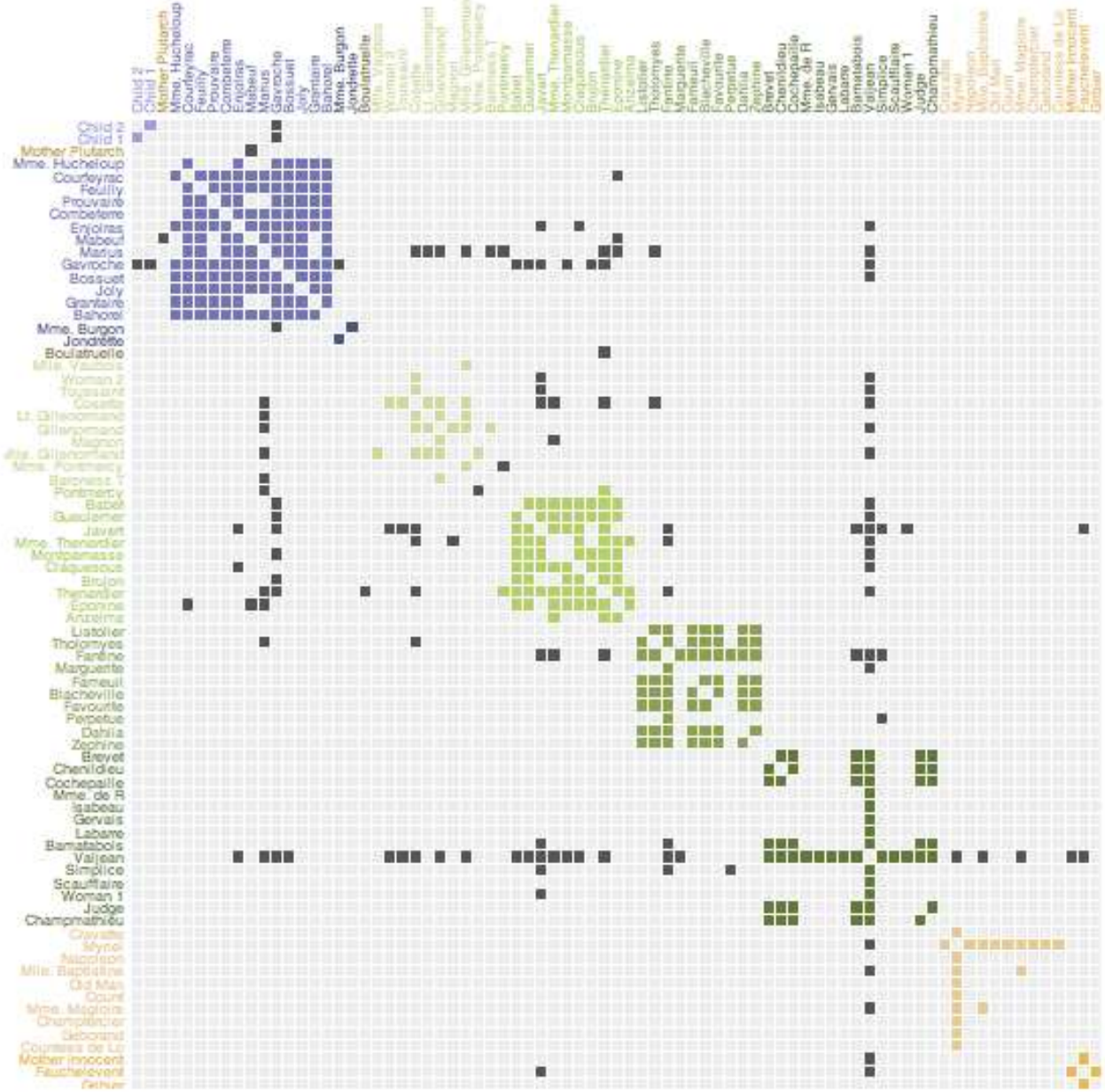
From: *The Eyes Have It:A Task by Data Type Taxonomy for Information Visualizations*

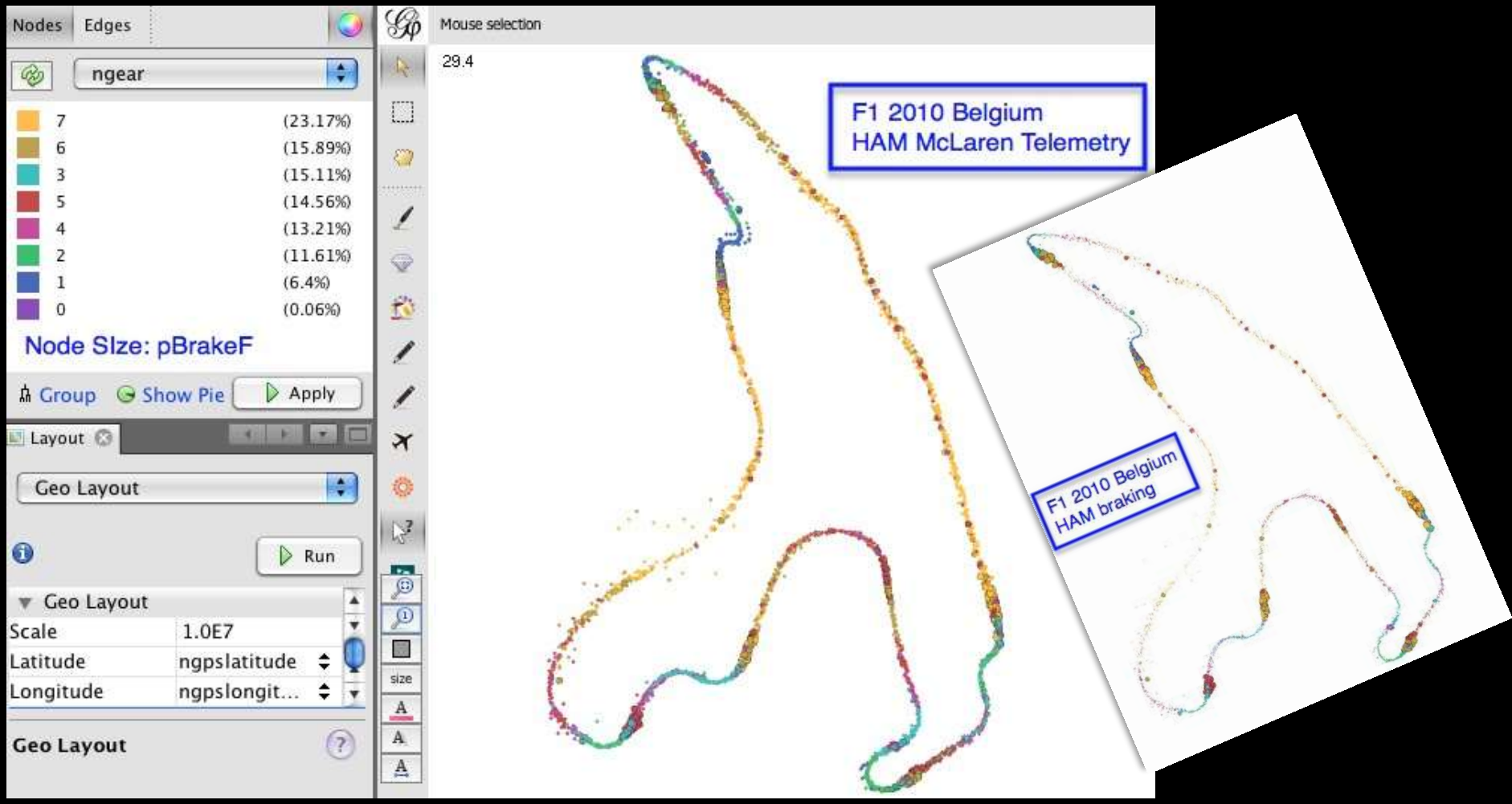

Ed-tech
IWMW
Libraries
Accessibility
Info pros
Spanish?
Museums
jamesclay
timbucktoo
jiscdigital
mweller
josiefraser
WordShore
psychemedia
andypowe11
Philbradley
ostephens
mariekeguy
paulwalk
research
inform
TwitPic
hashtags
lorcanD
PaulMiller
MrTweet
m1ke_ellis
MLA_gov
NickPoole1
Culture24

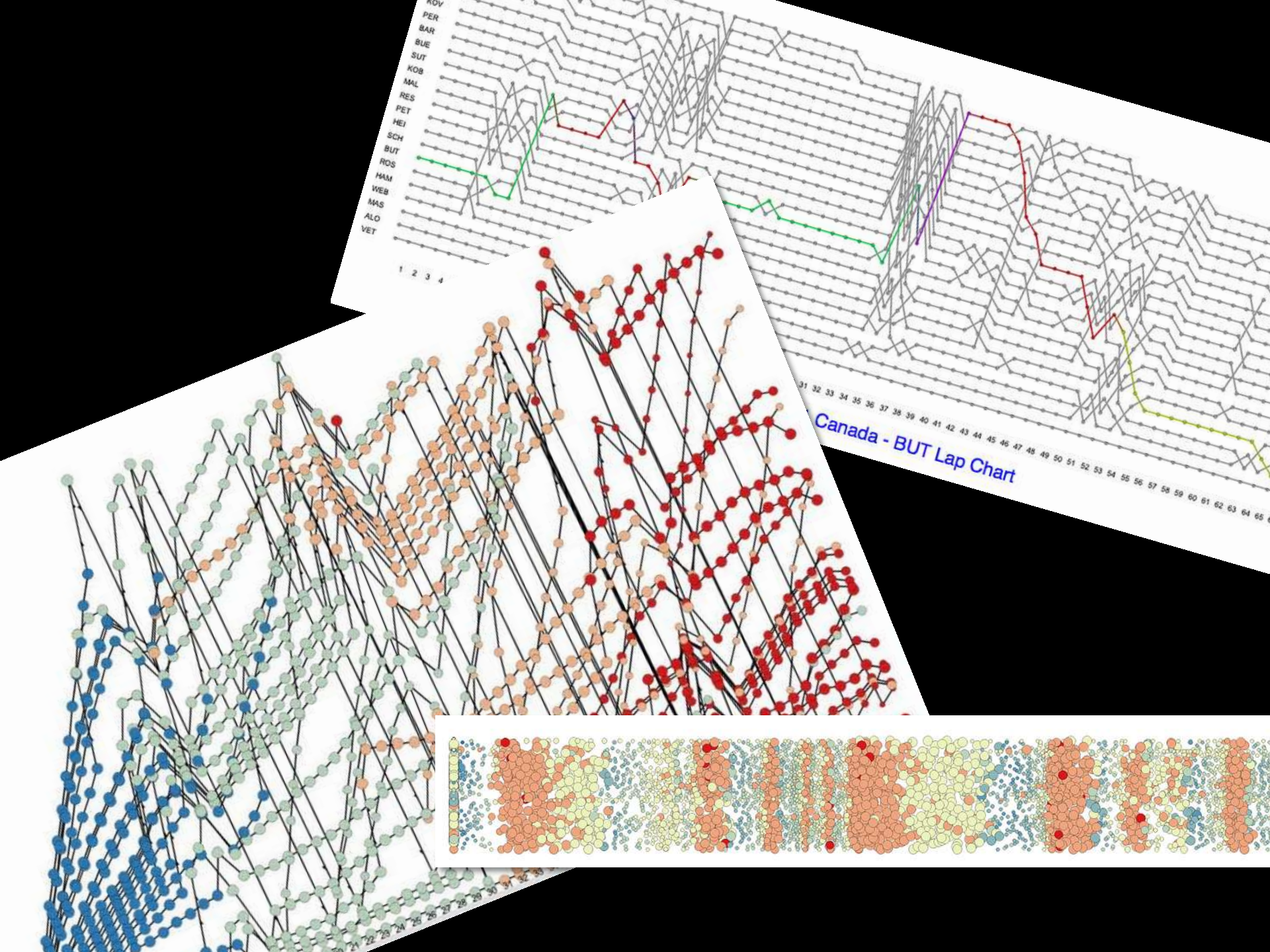

/catherinecronin/education-elearning
jiscoetis
JISC
andyjowe11
JISC_Techdis
jiscdigital
JISCLegal
JISCAdvance
netskills
jiscinfonet
JISC_RSC_YH


GrahamAttwell
mhawksey
zakmensah
dajbelshaw
jisccetis
andypowe11
shrifootring
gconole
ambrouk
jamesclay
timbuckteeth
KavuBob
mattpearson
billgx
markpower
JISC

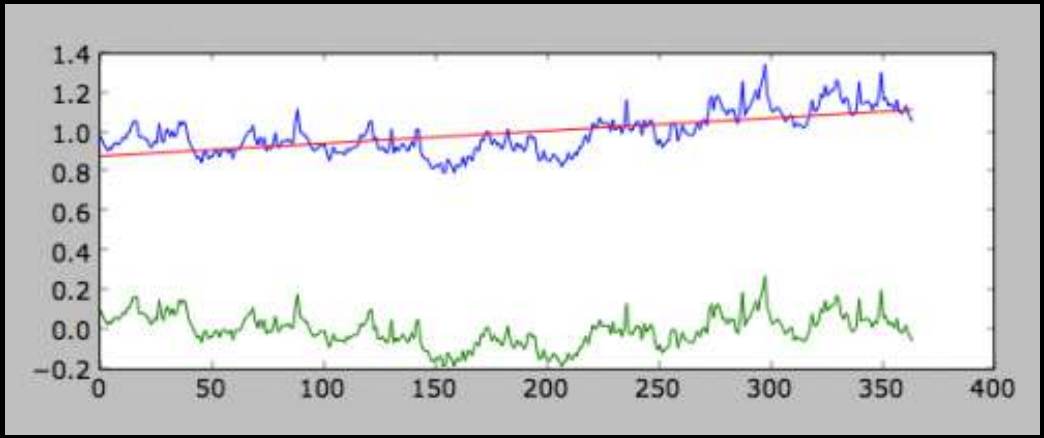
Matrix Diagrams

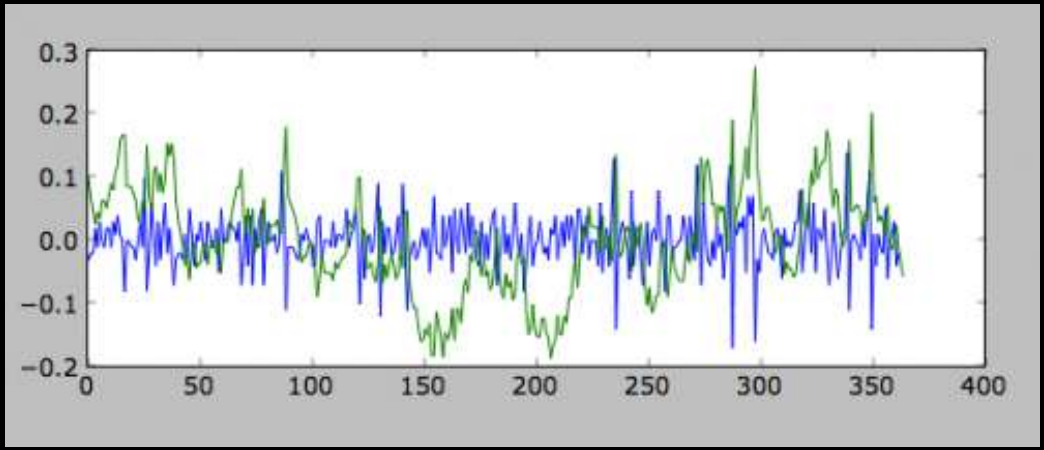


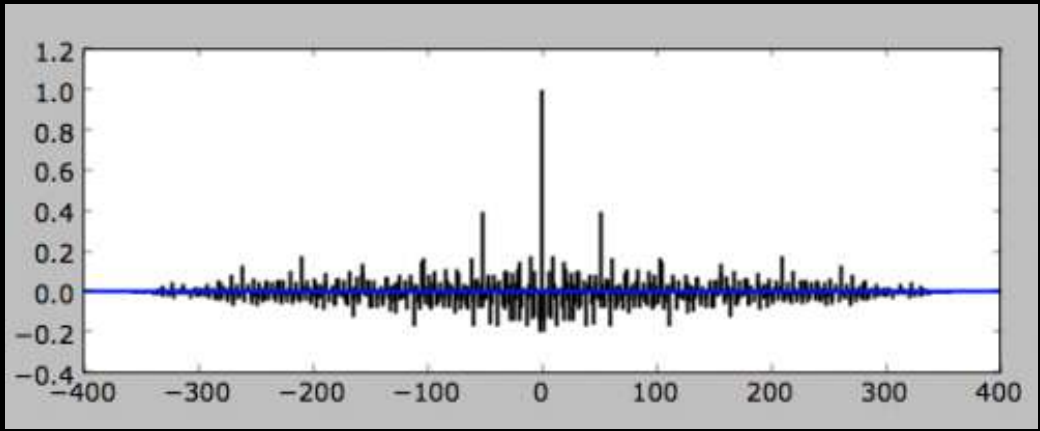

Nodes
Edges
ngear
7 (23.17%)
6 (15.89%)
3 (15.11%)
5 (14.56%)
4 (13.21%)
2 (11.61%)
1 (6.4%)
0 (0.06%)
Node SIze: pBrakeF
Group
Show Pie
Apply
Layout
Geo Layout
Run
Geo Layout
Scale 1.0E7
Latitude ngpslatitude
Longitude ngpslongit...
Geo Layout
Mouse selection
29.4
size
F1 2010 Belgium
HAM McLaren Telemetry
F1 2010 Belgium
HAM braking


KOV
PER
BAR
BUE
SUT
KOB
MAL
RES
PET
HEI
SCH
BUT
ROS
HAM
WEB
MAS
ALO
VET
Canada - BUT Lap Chart

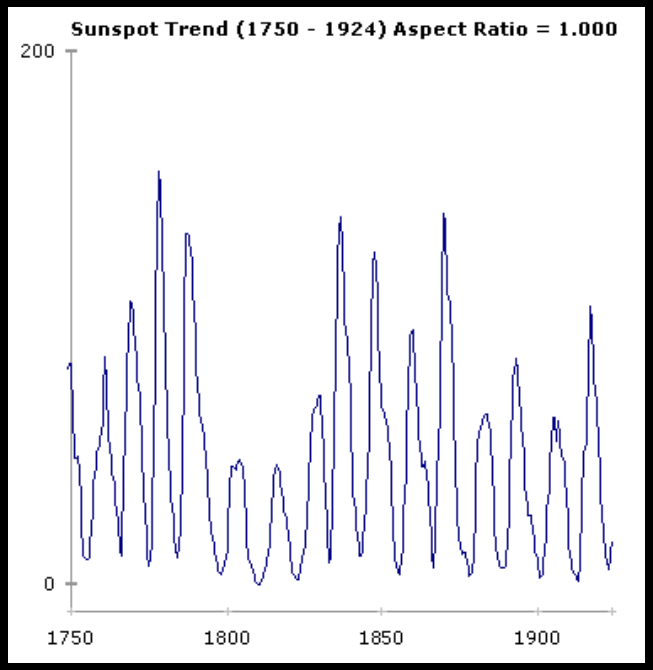
Time Series Data


1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
50
100
150
200
250
300
350
400


0.3
0.2
0.1
0.0
−0.1
−0.2
0
50
100
150
200
250
300
350
400


1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−400
−300
−200
−100
0
100
200
300
400

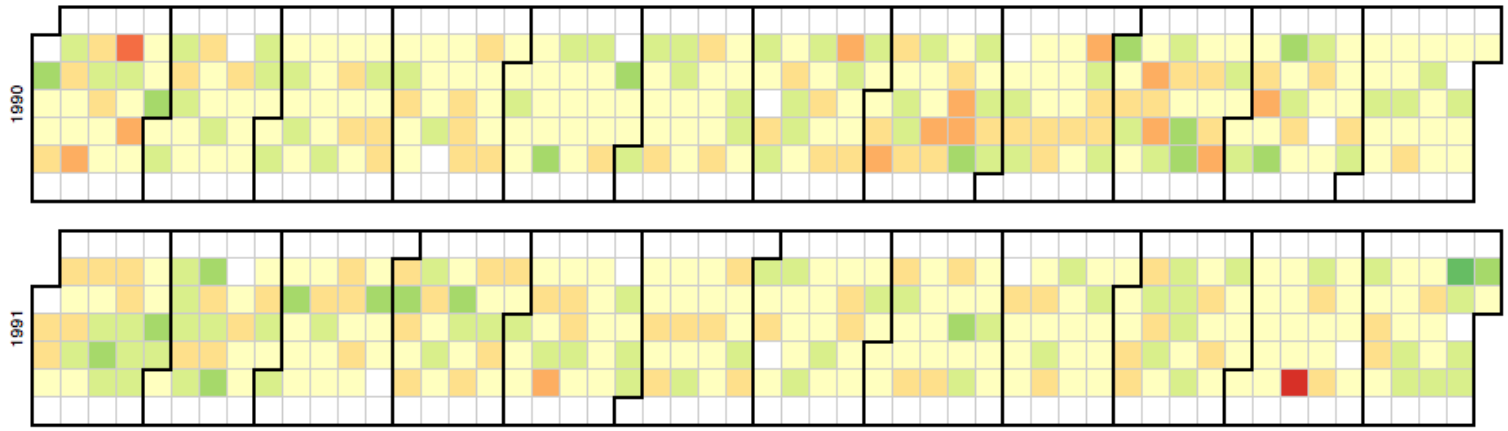
“Banking to 45 degrees” (Cleveland)



“The aspect ratio is vital because it has a large impact on our ability to judge rate of change. A number of studies in visual perception have shown that our ability to judge the relative slopes of line segments on a graph is maximized when the absolute values of the orientations of the segments are centered on 45 degrees.”

d3.js

Calendar View



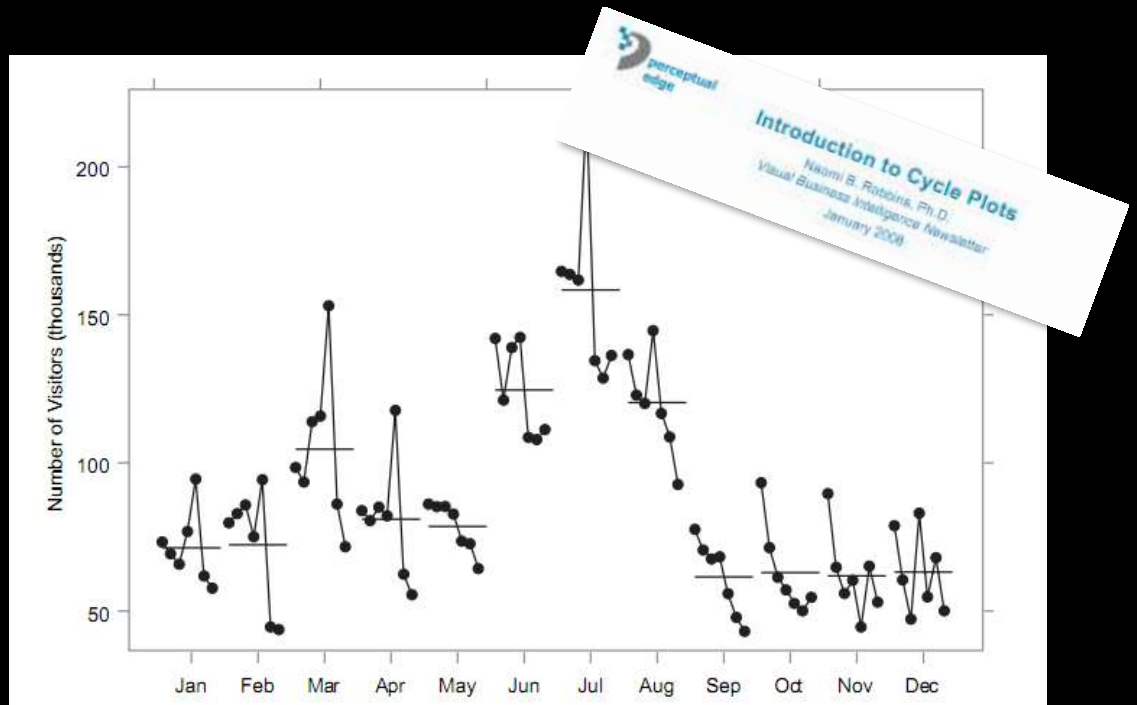
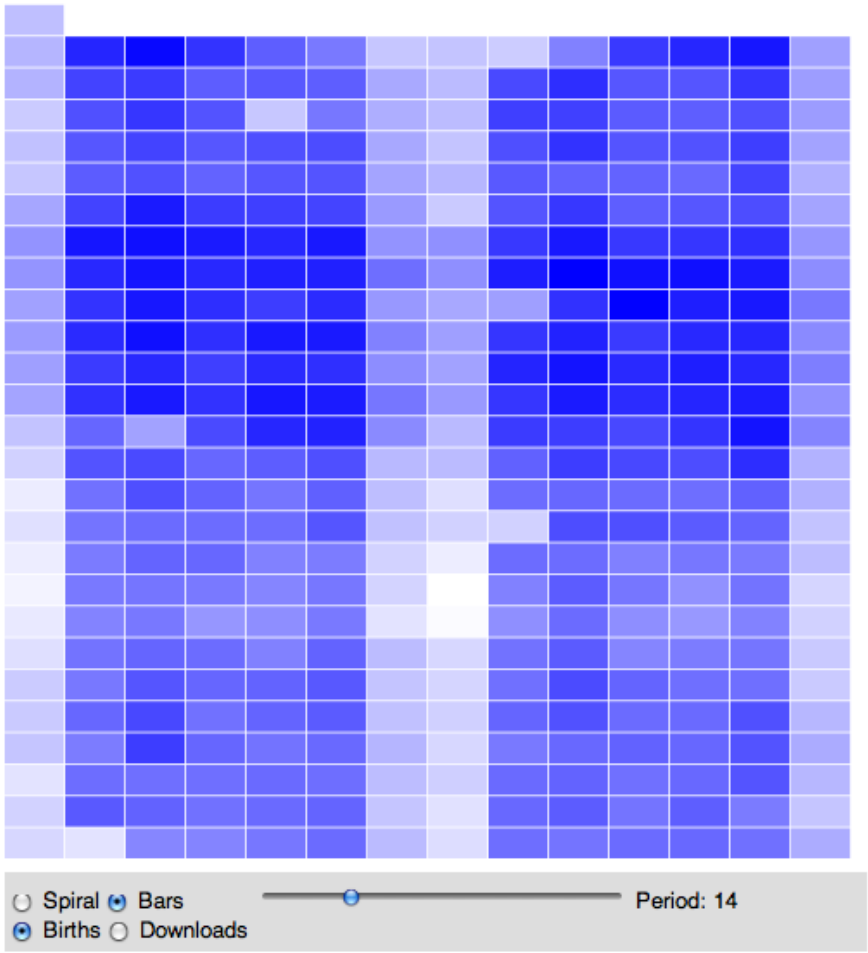


Figure 5: This cycle plot makes it much easier to visualize the trend of a given month over the seven years.

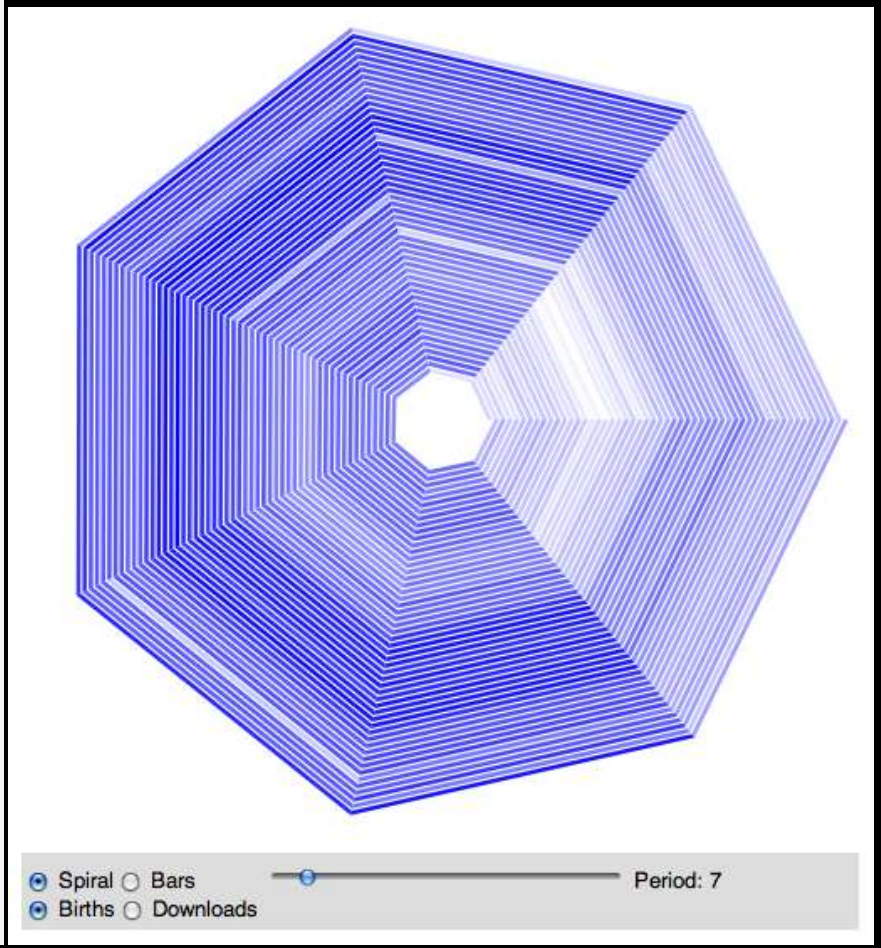
Interactive Version

Below is an interactive version of a spiral display (implemented using Protovis), so you can try it out and draw your own conclusions. This requires a current version of Safari, Chrome, or FireFox (this should work in Internet Explorer 9 in principle, but for some reason doesn't).

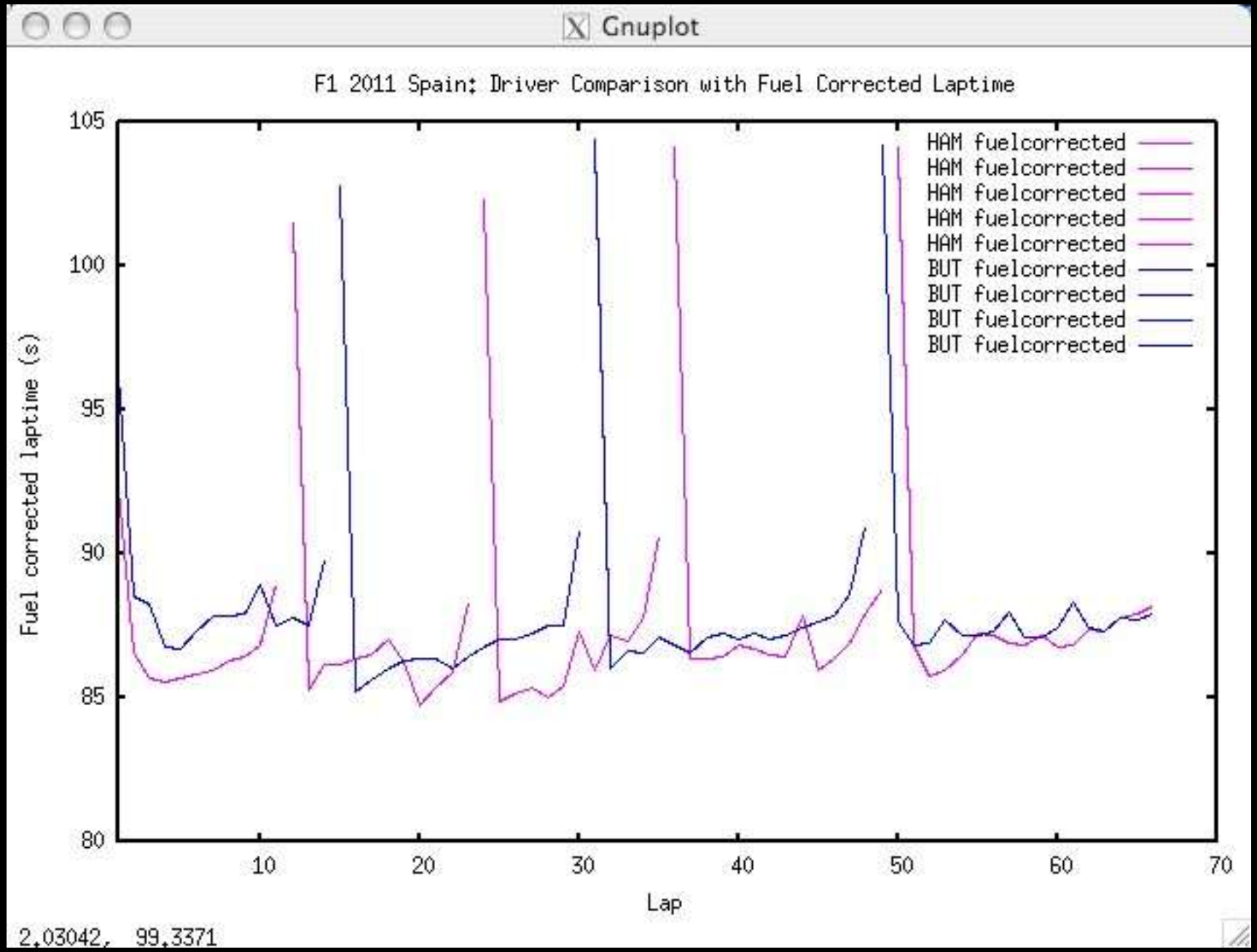
Move the slider to control the period of the spiral. Change the display type between spiral and bars with the radio buttons on the left.

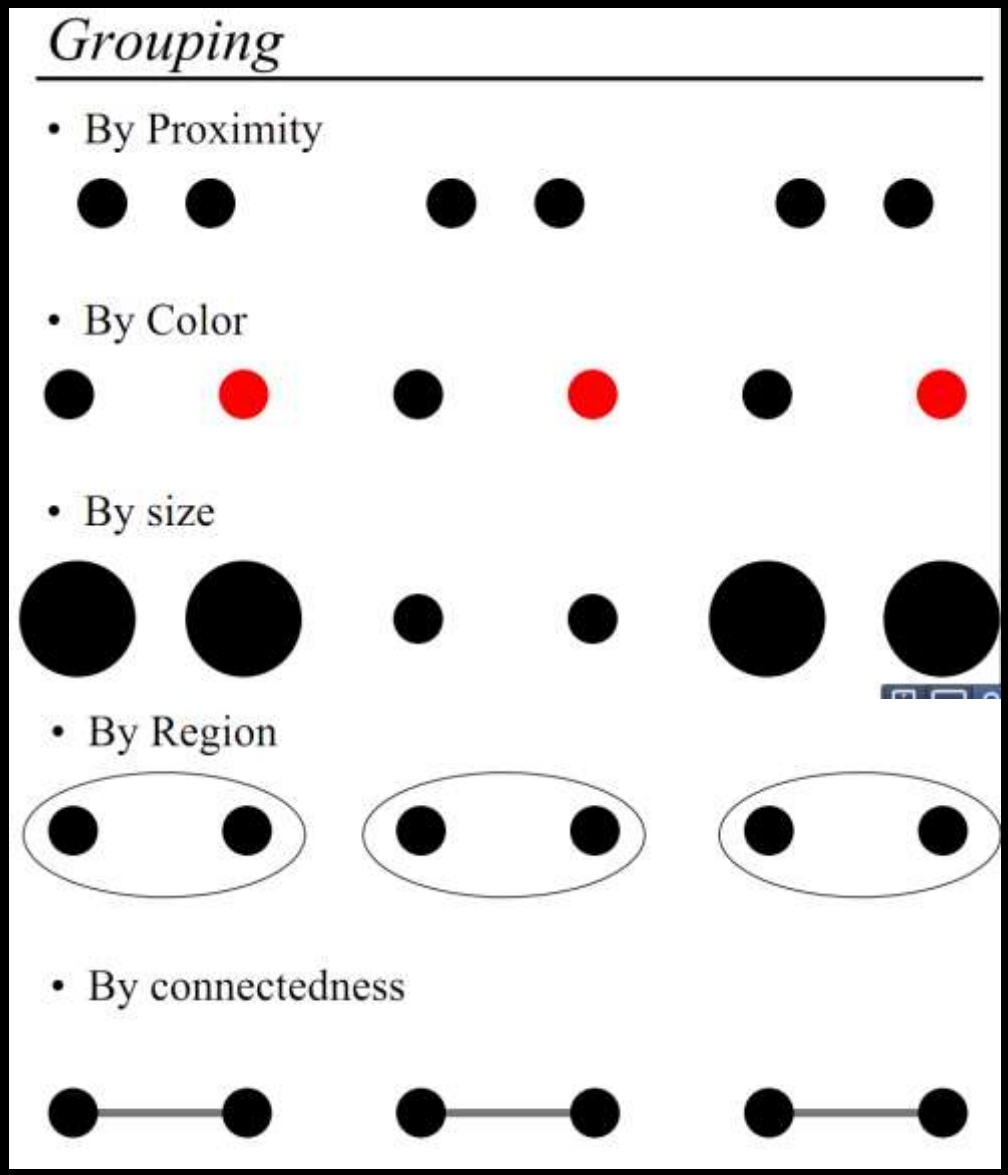


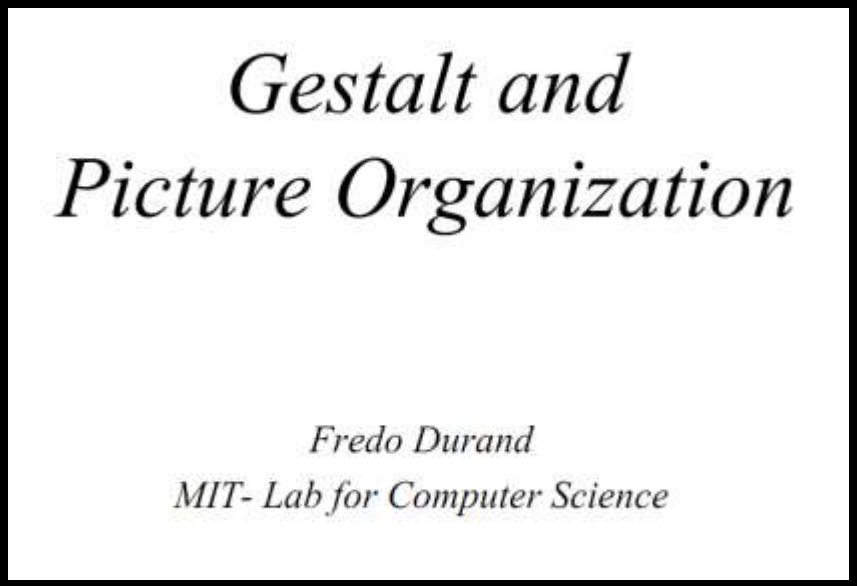
http://eagereyes.org/techniques/spirals

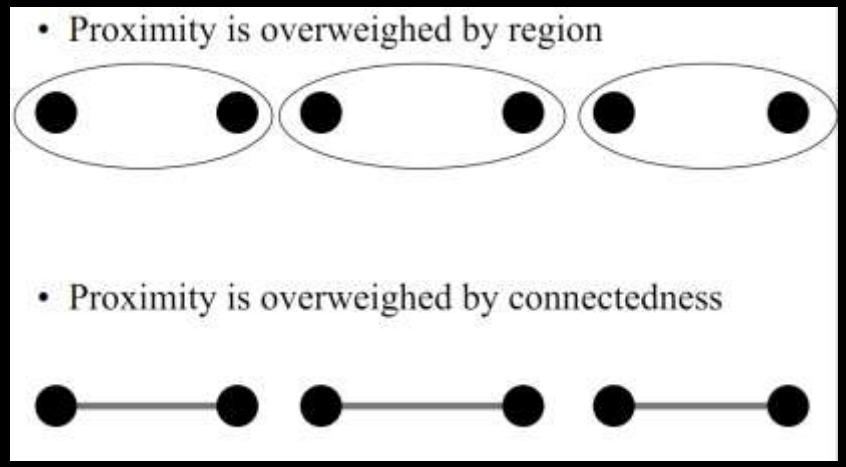


Gestalt Theory of Visual Perception

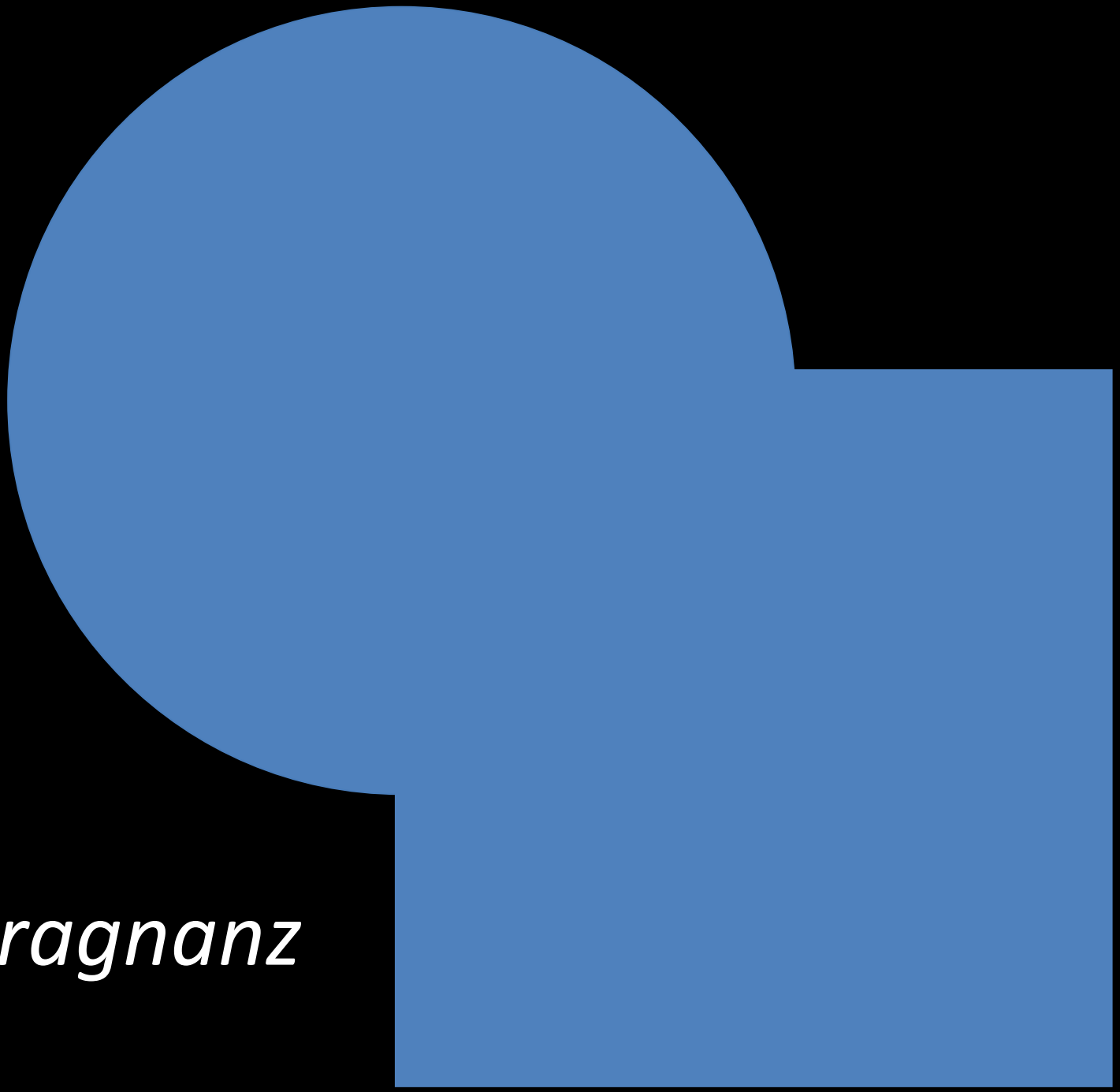

Gnuplot
F1 2011 Spain: Driver Comparison with Fuel Corrected Laptime
HAM fuelcorrected
HAM fuelcorrected
HAM fuelcorrected
HAM fuelcorrected
HAM fuelcorrected
BUT fuelcorrected
BUT fuelcorrected
BUT fuelcorrected
BUT fuelcorrected
Fuel corrected laptime (s)
105
100
95
90
85
80
10
20
30
40
50
60
70
Lap
2.03042, 99.3371



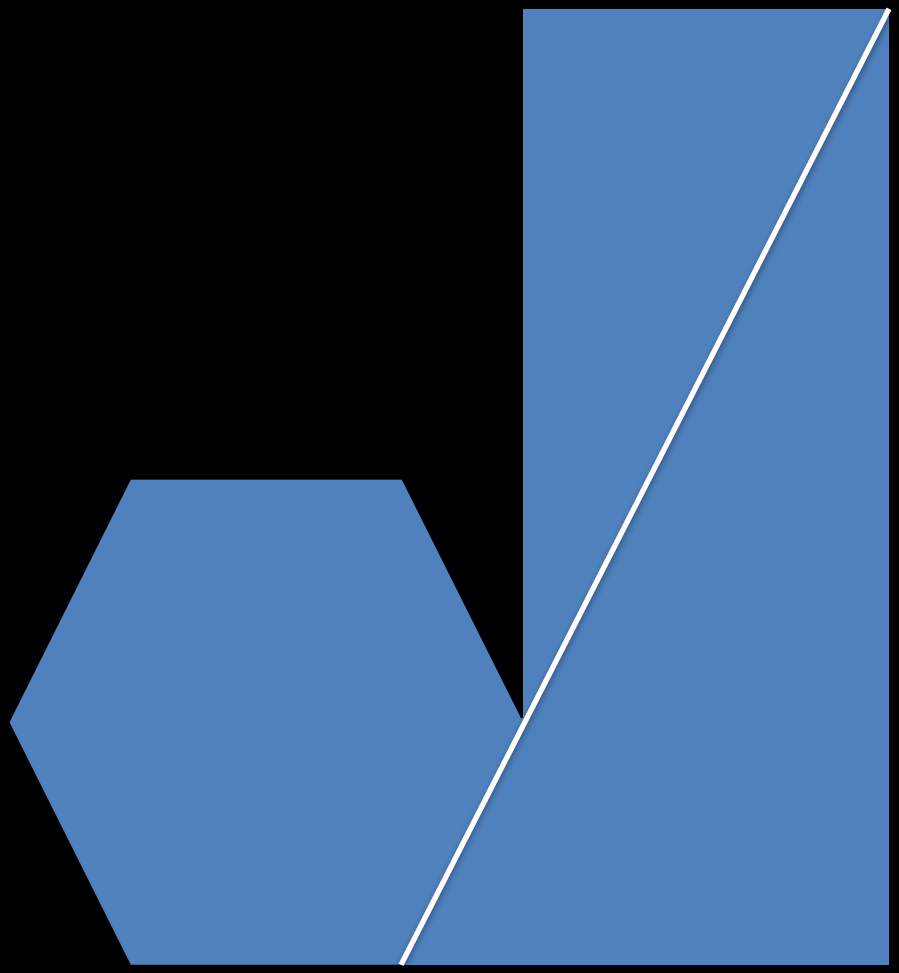




http://hci.stanford.edu/courses/cs448b/papers/Durand_siggraph_Gestalt_talk.pdf



Pragnanz



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox

Print/export

Structural information theory

From Wikipedia, the free encyclopedia

Structural information theory (**SIT**) is a theory about human perception and in particular about perceptual organization: the way the human visual system organizes a raw visual stimulus into objects and object parts. SIT was initiated, in the 1960s, by Emanuel Leeuwenberg[1][2][3] and has been developed further by Hans Buffart, Peter van der Helm, and Rob van Lier. It has been applied to a wide range of research topics, mostly in visual form perception but also in, for instance, visual ergonomics, data visualization, and music perception.

SIT began as a quantitative model of visual pattern classification. Nowadays, it also includes quantitative models of symmetry perception and amodal completion, and it is theoretically founded in formalizations of visual regularity and viewpoint dependency. SIT has been argued to be the best defined and most successful extension of Gestalt ideas.[4] It is the only Gestalt approach providing a formal calculus that generates plausible perceptual interpretations.

Contents [hide]

The simplicity principle

[edit]

Although visual stimuli are fundamentally multi-interpretable, the human visual system usually has a clear preference for only one interpretation. To explain this preference, SIT introduced a formal coding model starting from the assumption that the perceptually preferred interpretation of a stimulus is the one with the simplest code. A simplest code is a code with minimum information load, that is, a code that enables a reconstruction of the stimulus using a minimum number of descriptive parameters. Such a code is obtained by capturing a maximum amount of visual regularity and yields a hierarchical organization of the stimulus in terms of wholes and parts.

The assumption that the visual system prefers simplest interpretations is called the simplicity principle.[5] Historically, the simplicity principle is an information-theoretical descendant of the Gestalt law of Prägnanz,[6] which was based on the natural tendency of physical systems to settle into stable minimum-energy states. Furthermore, just as the later-proposed minimum description length principle in algorithmic information theory (AIT), it can be seen as a formalization of Occam's Razor in which the best hypothesis for a given set of data is the one that leads to the largest compression of the data.

To what extent does the viewer use the visualisation to inform the creation of a model that they then interpret in order to spot the differences that make a difference in the visualisation?

Information required to generate a visualisation

VS

Information revealed by a visualisation

(Probably no time for)

QUESTIONS…?

http://blog.ouseful.info
@psychemedia

"Mandelbrot Set Fractal - Milky Way"
- dominicspics

"Every block of stone has a statue inside it and it is the task of the sculptor to discover it.”

- Michelangelo