

notes on visualization
OF HIGHLY COMPLEX DATA
santiago v lombeyda

santiago v lombeyda

data



new questions
~~*questions*~~

analytical
👉 answers

visual inspection
👉 answers

what includes visualization

data



phantom omni

more und



entation



Acquire

Obtain the data, whether from a file on a disk or a source over a network.

Parse

Provide some structure for the data's meaning, and order it into categories.

Filter

Remove all but the data of interest.

Mine

Apply methods from statistics or data mining as a way to discern patterns or place data in mathematical context.

Represent

Choose a basic visual model, such as a bar graph, list, or tree.

Refine

Improve the basic representation to make it clearer and more visually engaging.

Interact

Add methods for manipulating the data or controlling what features are visible.

Acquire

Obtain the data, whether from a file on a disk or a source over a network.

Pa

Pr

da

in

exploratory phase



```
graph LR; Acquire --> Parse; Parse --> Filter; Filter --> Mine; Mine --> Represent; Represent --> Refine; Refine --> Interact; Interact --> Acquire
```

Acquire

Obtain the data, whether from a file on a disk or a source over a network.

Parse

Provide some structure for the data's meaning, and order it into categories.

Filter

Remove all but the data of interest.

Mine

Apply methods from statistics or data mining as a way to discern patterns or place data in mathematical context.

Represent

Choose a basic visual model, such as a bar graph, list, or tree.

Refine

Improve the basic representation to make it clearer and more visually engaging.

Interact

Add methods for manipulating the data or controlling what features are visible.

multivariable visualization...

how many variables can you visualize?

as many as you want!

how many variables can you understand?

3? 4?

Collin Ware study:

encoding 4 variables

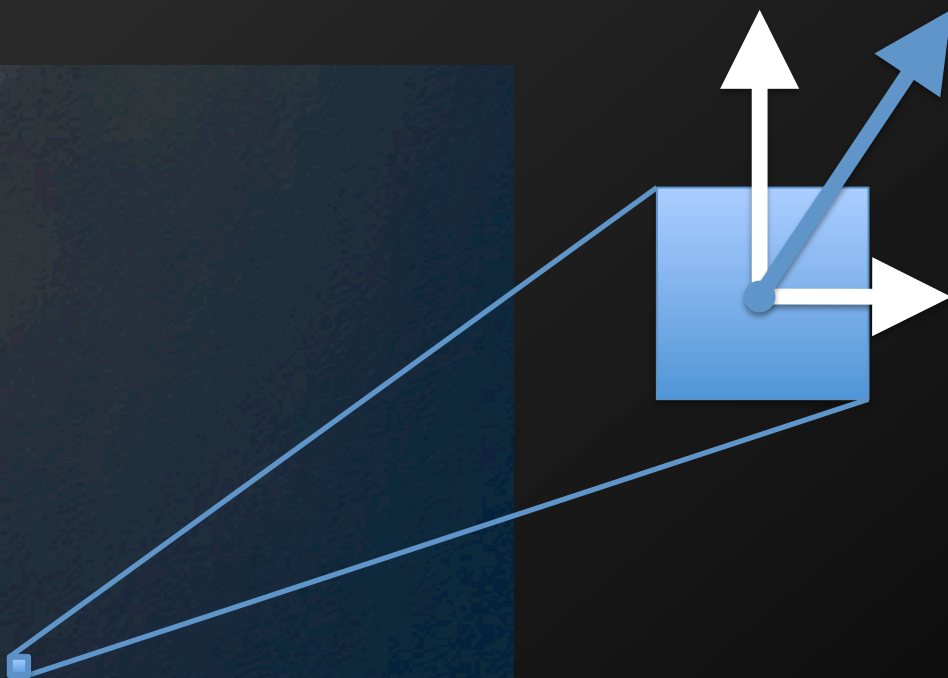
i.e. encoding 2 variables on map

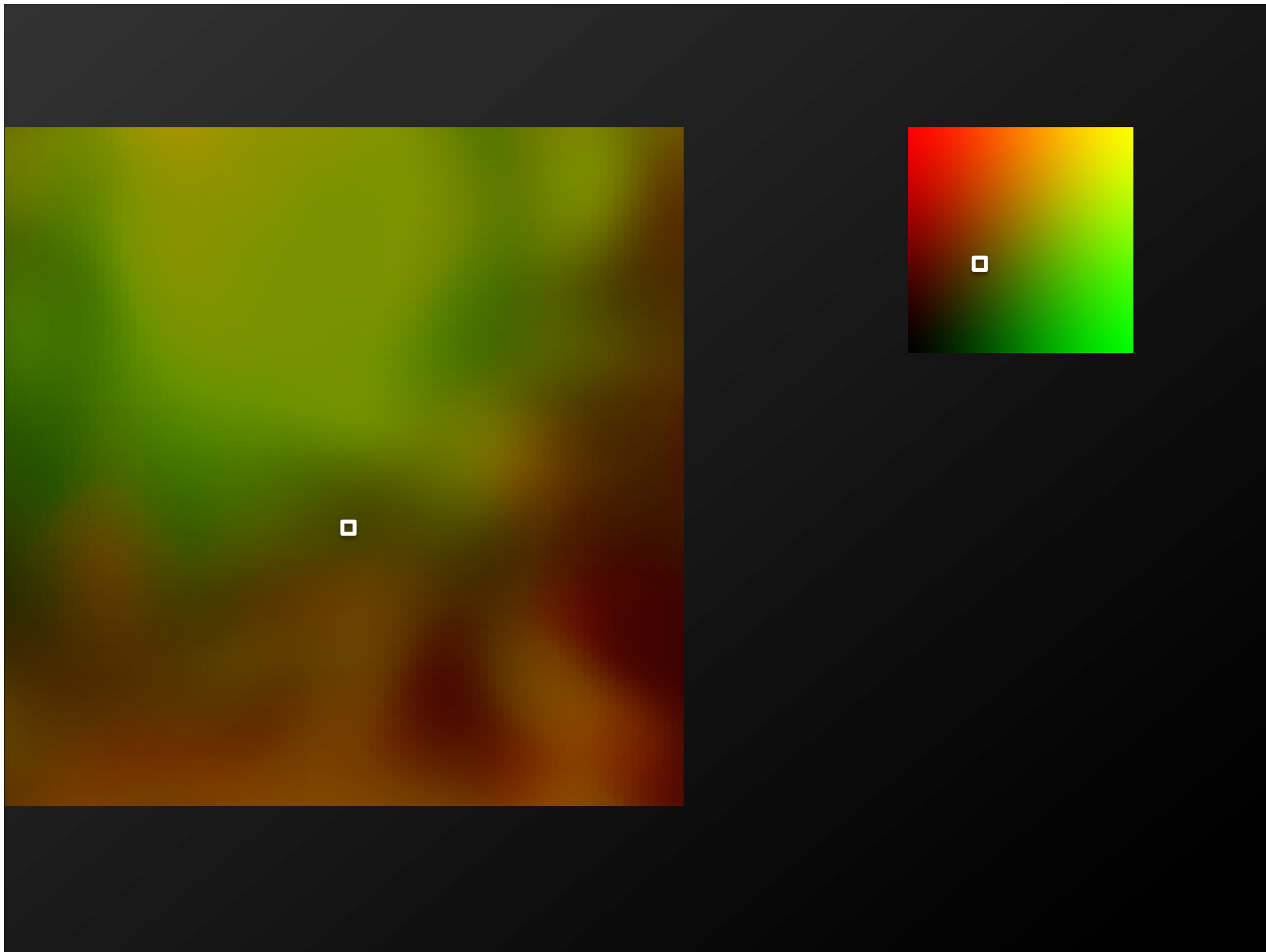
(x, y, u, v)

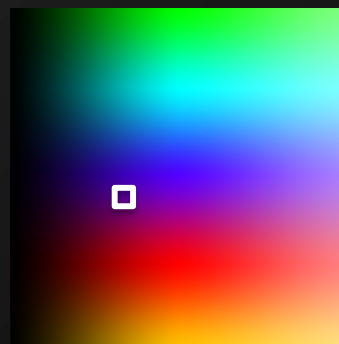
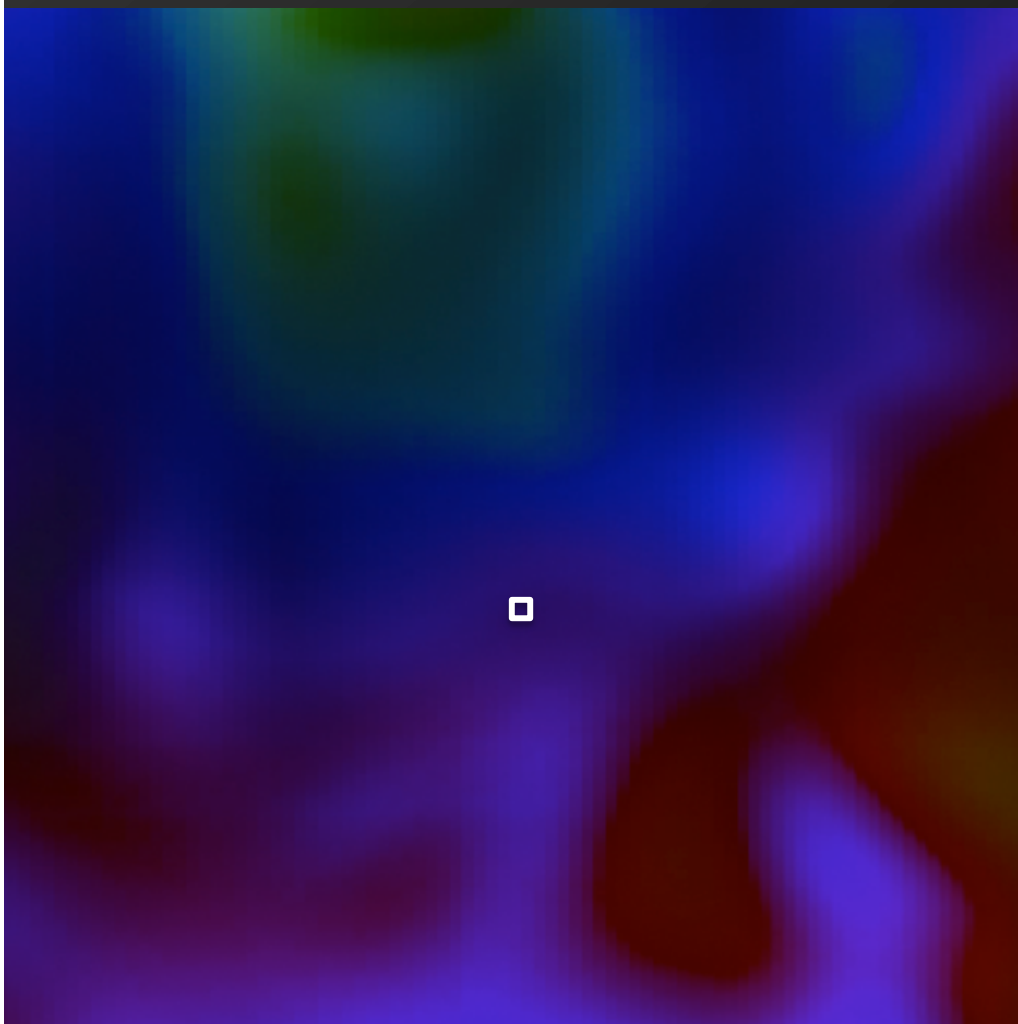
Qton:

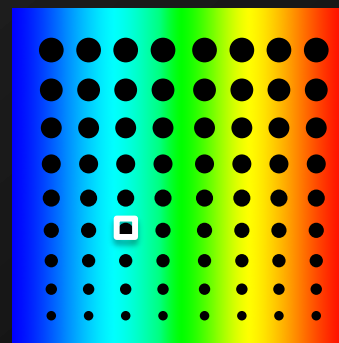
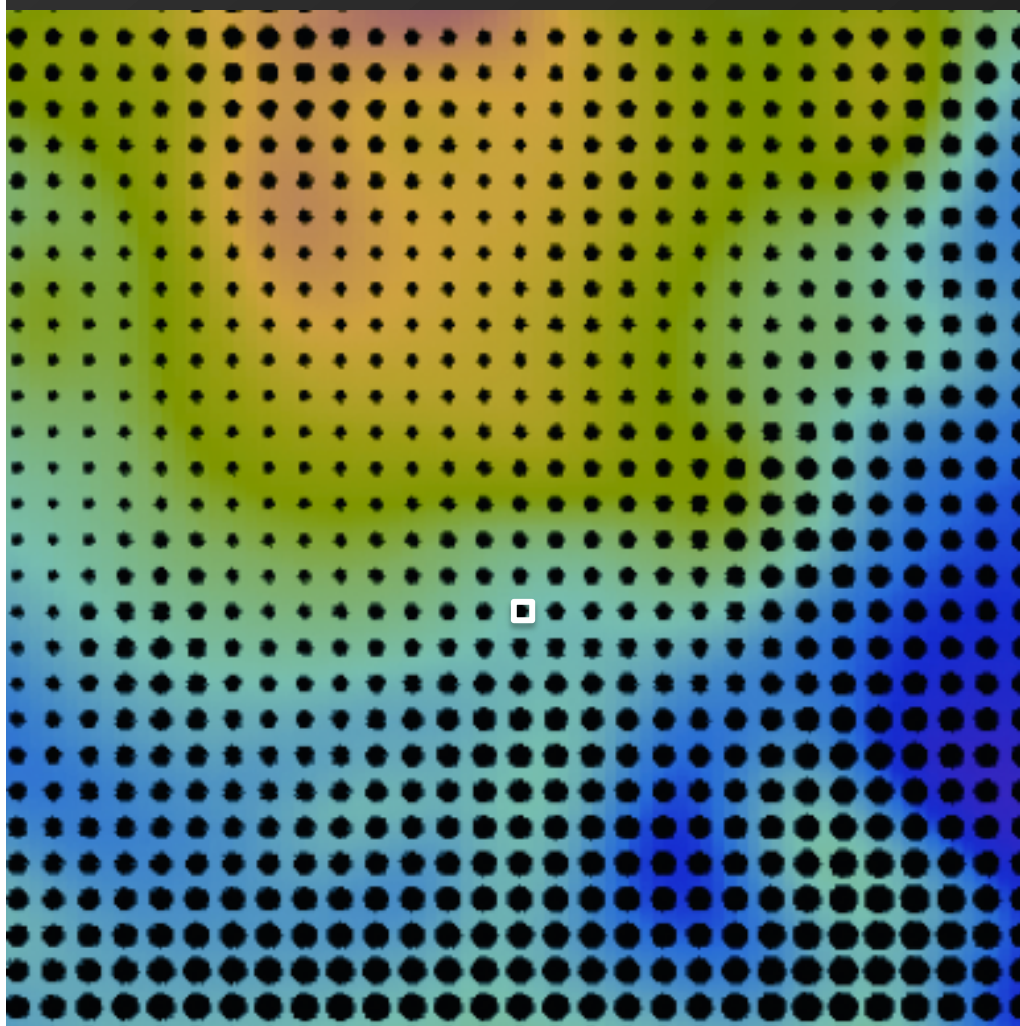
putative units of pre-attentive human texture perception,
analogous to a phoneme in speech recognition

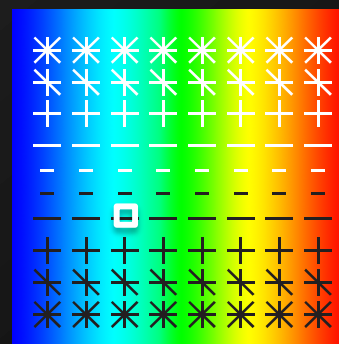
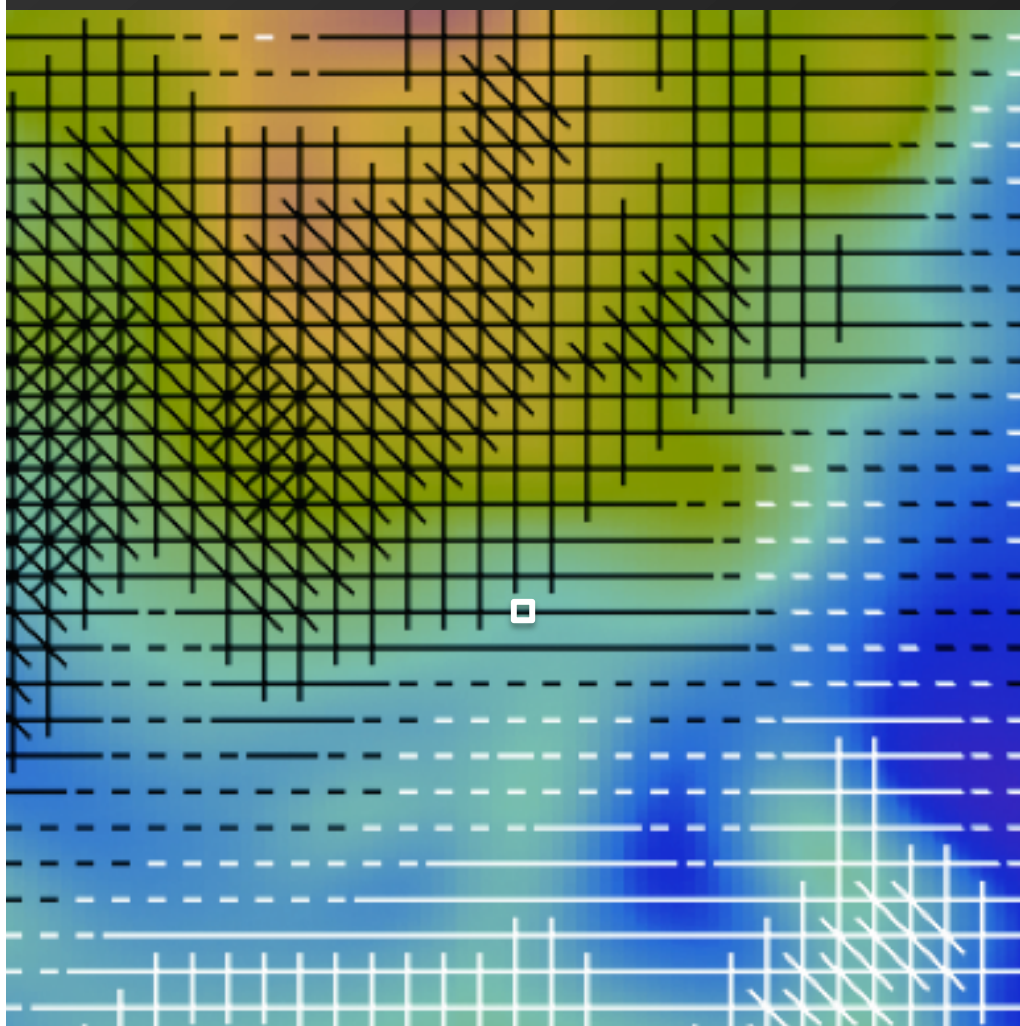
(color salad, hold the qtons!)

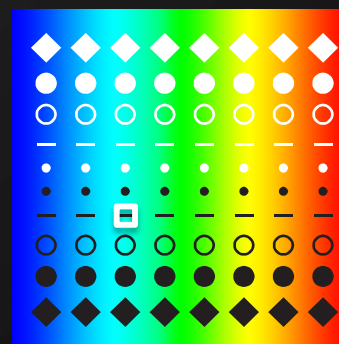
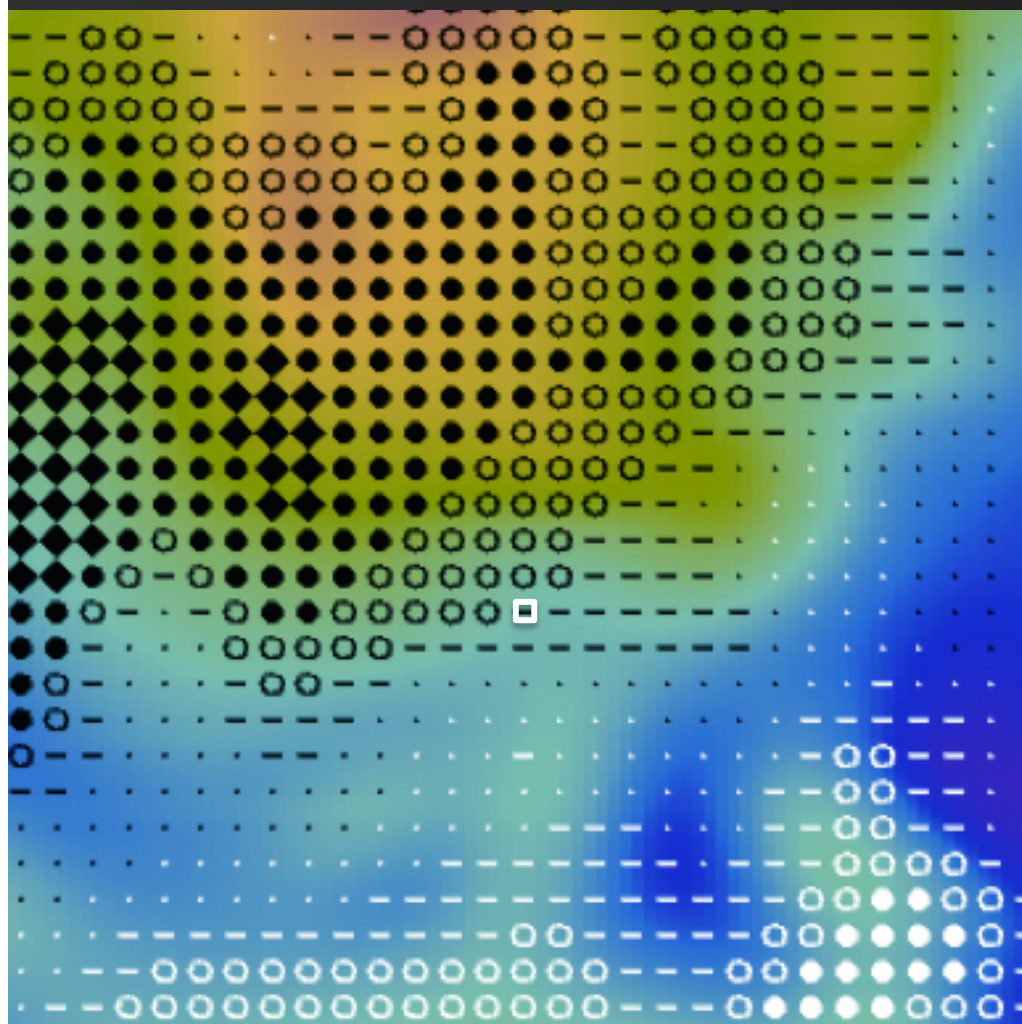




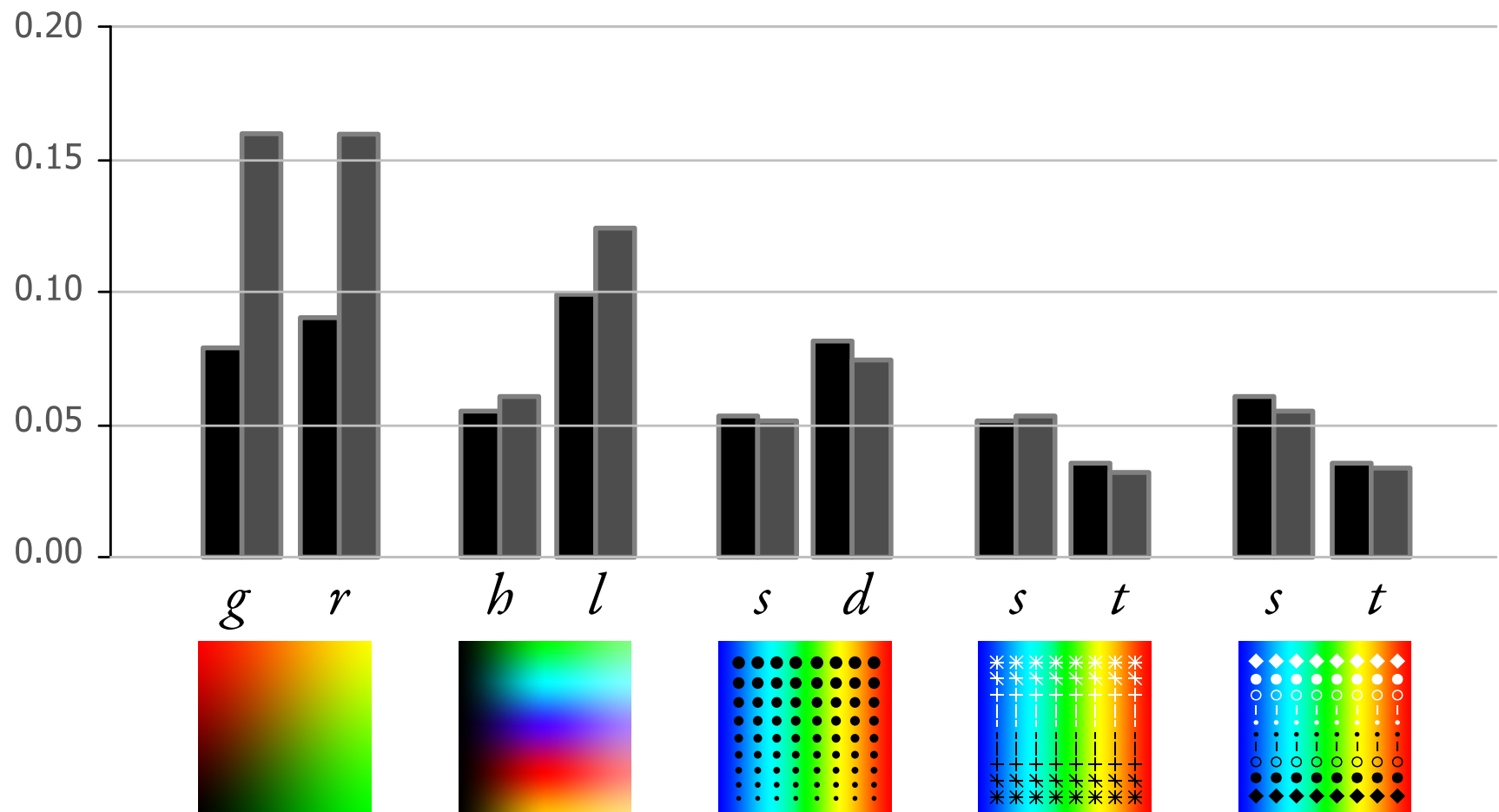








error: integral  separable 





reading order/focal flow



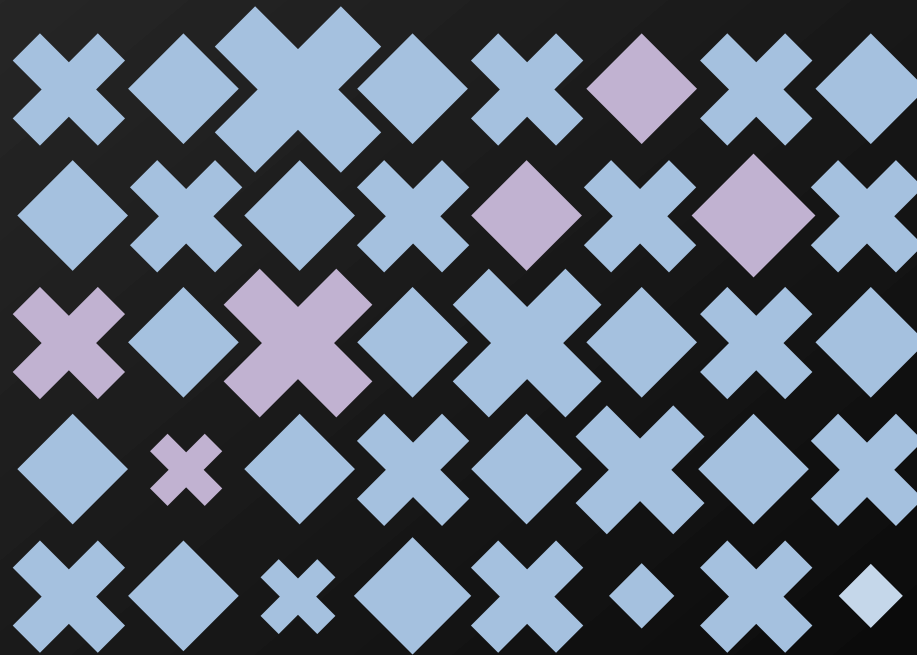
what do you see first? second?

why is it important?

what does unknown order create?







science
(data analysis, visual analytics)



graphics



graphic design/art
(human perception, aesthetics)

visualization

multivariable visualization...

how many variables can you visualize?

as many as you want!

how many variables can you understand?

simultaneous VS layered understanding

can we expand qtons design for further variables?

why?

task: multivariate scatterplots

issue: occlusion/overlapping

design idea: 3d qtons/glyphs

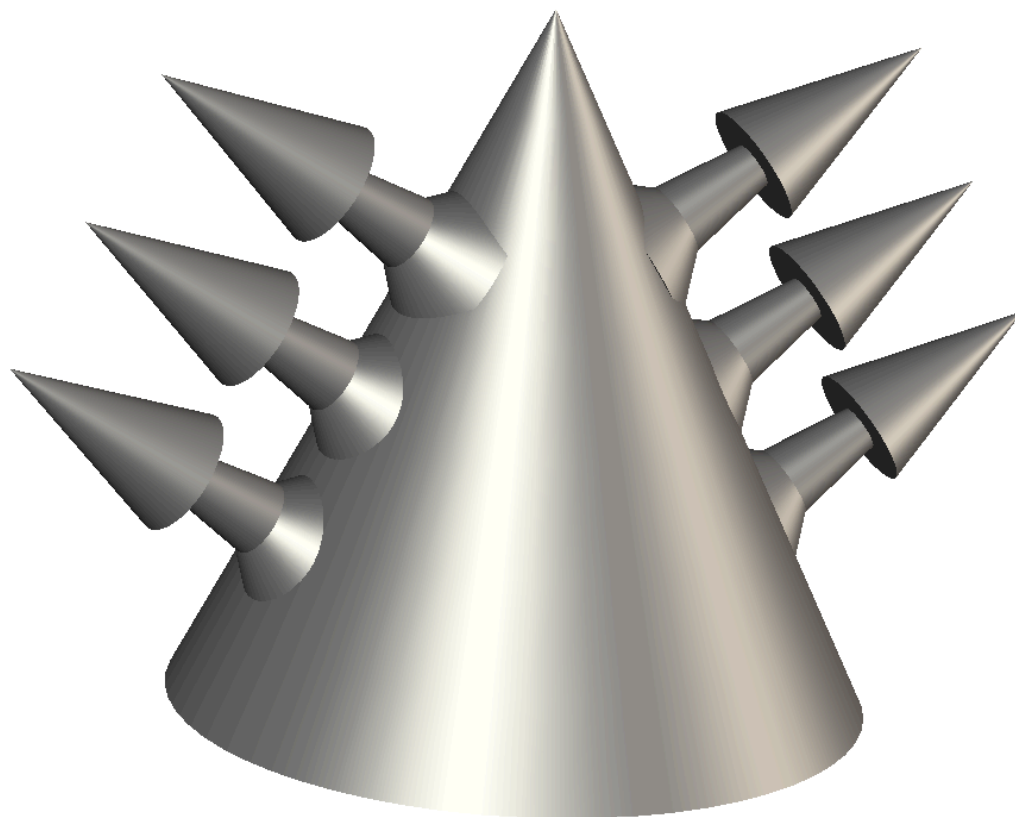
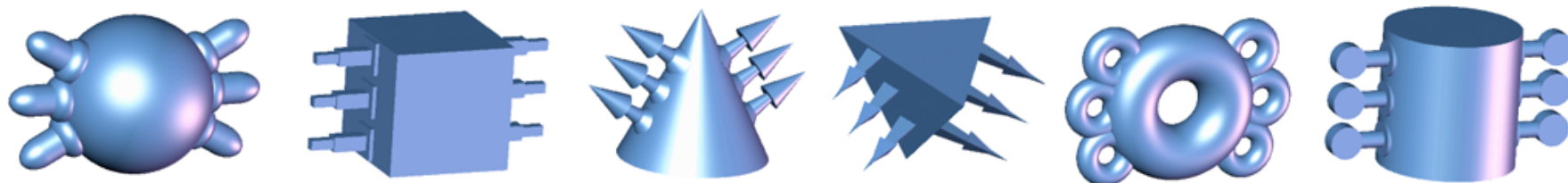
why 3d?

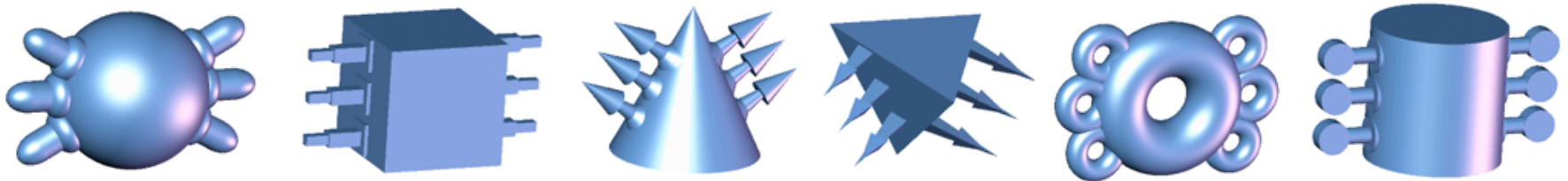
screens: 2d & paper: 2d

...but 3d cues:

occlusion / perspective / movement / shadows / stereoscopy

(issue: expected/acceptable/perceivable size)





(read order)

(attribute)

1,2

position (x,y)

3

shape

4

hue

5

left features

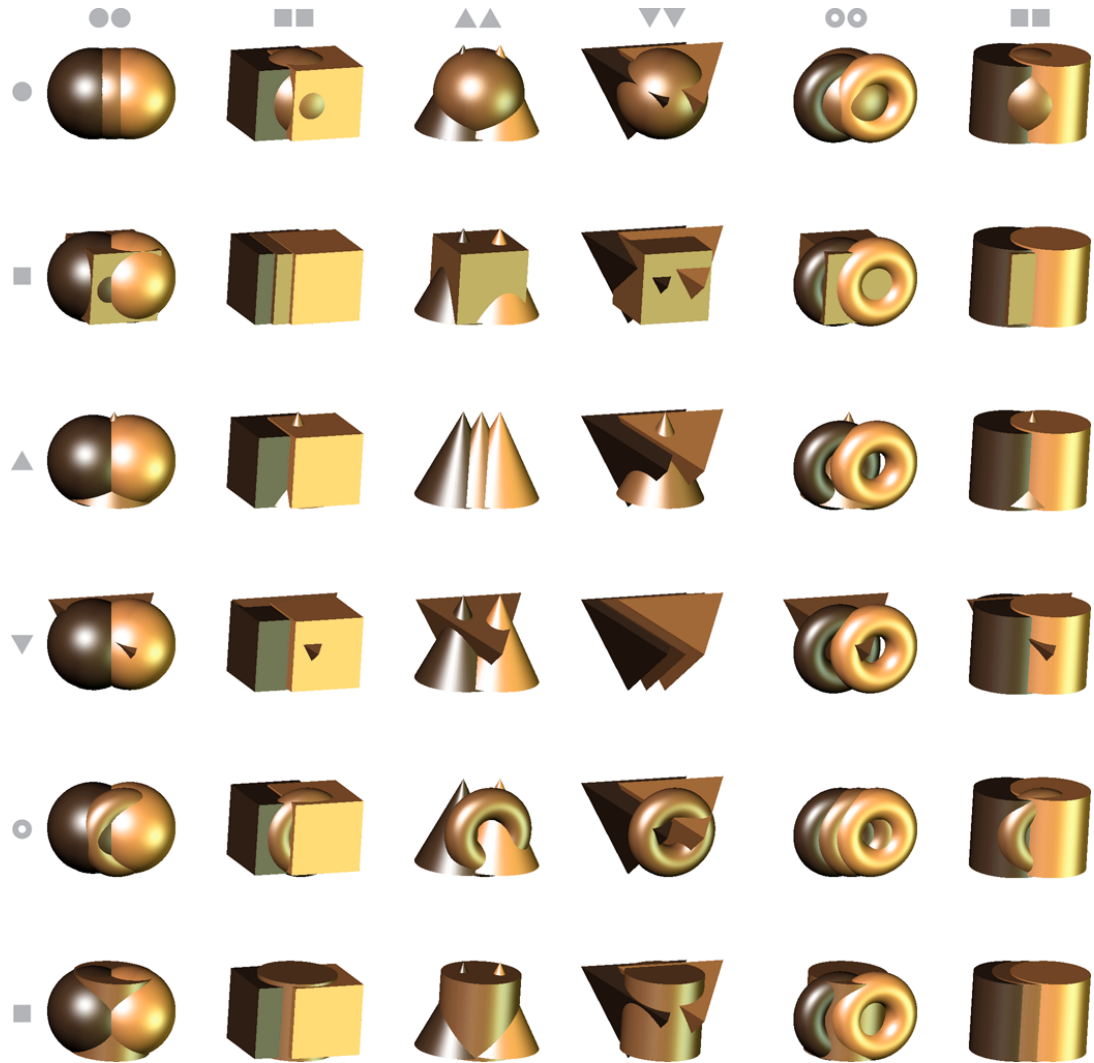
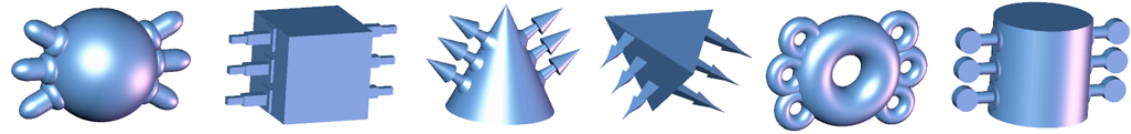
6

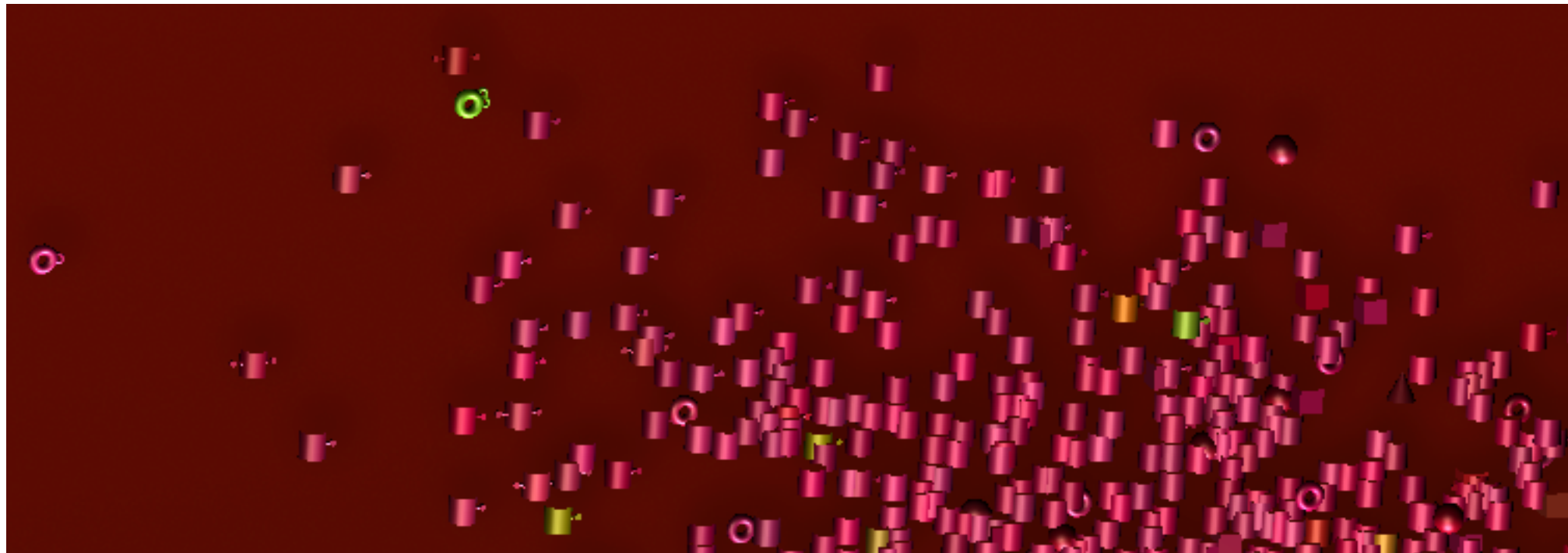
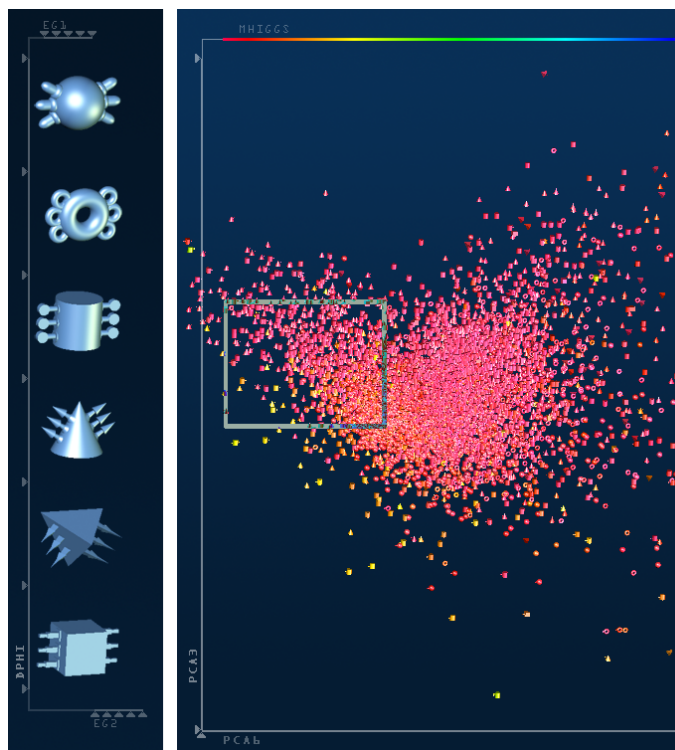
right features }

...

vibration, sound, etc...







I.

task must drive visualization design

II.

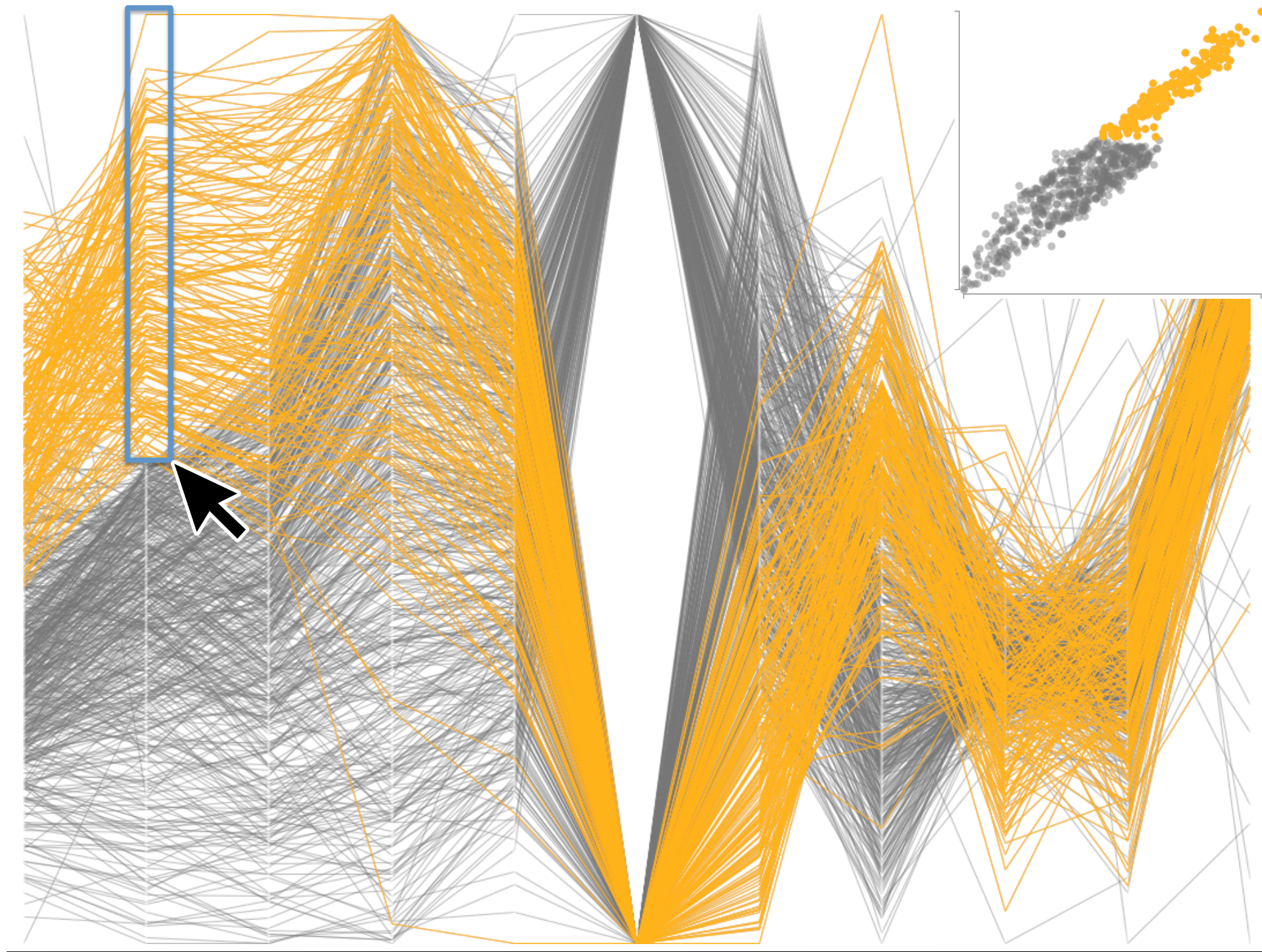
most tools out there will do *something*

III.

complement, cross reference, interact, ...

visualization representations
(from available tools, or tailored)

+ *interactivity*
(subsetting, queering, brushing)



visualization representations

(from available tools, or tailored)

+ *interactivity*

(subsetting, queering, brushing)

+ *data analysis*

(principal component analysis, clustering,...)

principal component analysis:

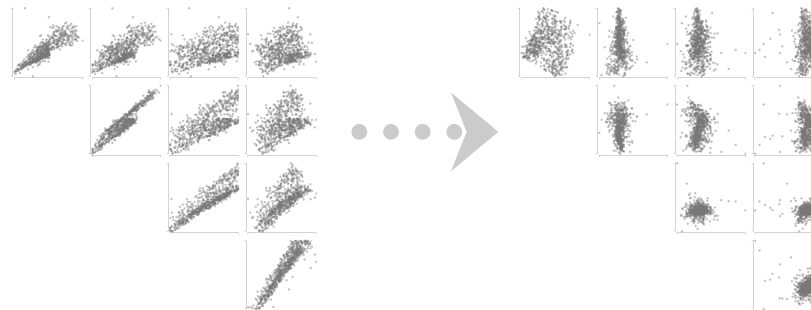
pca = data transformation

n “correlated” variables

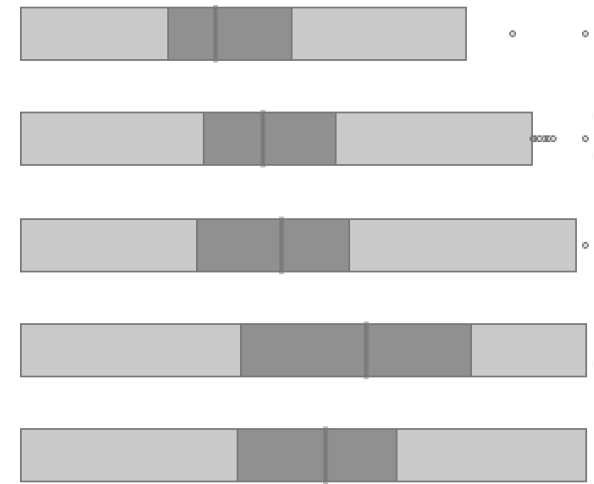
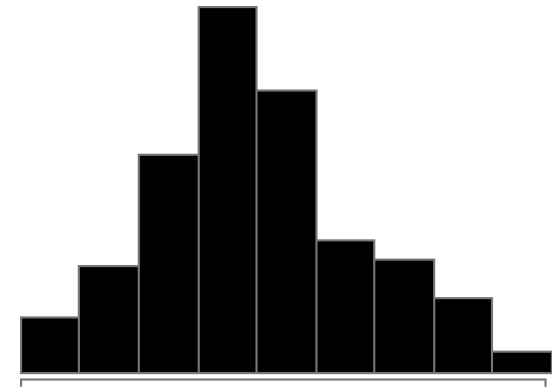
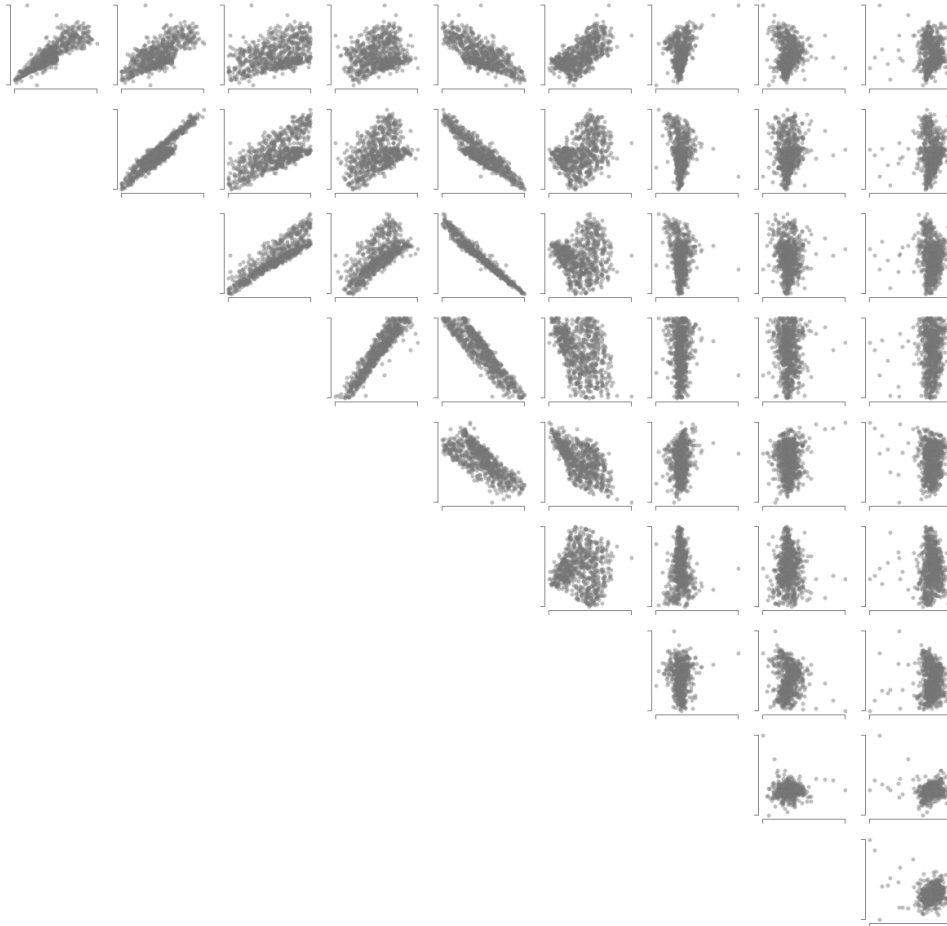
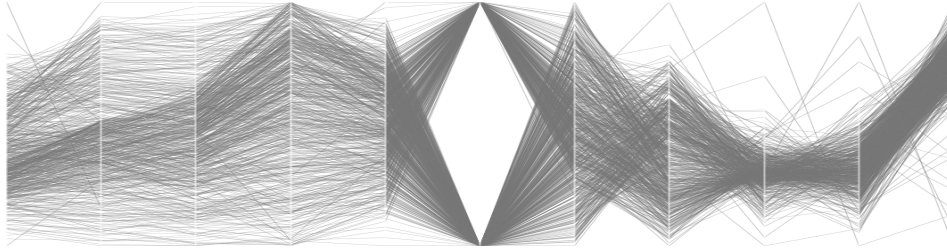


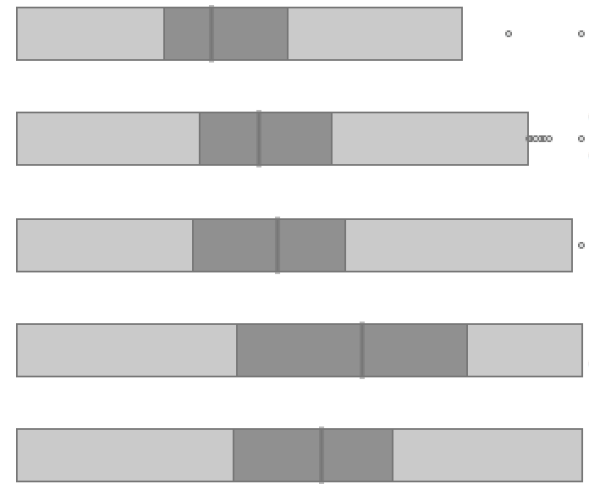
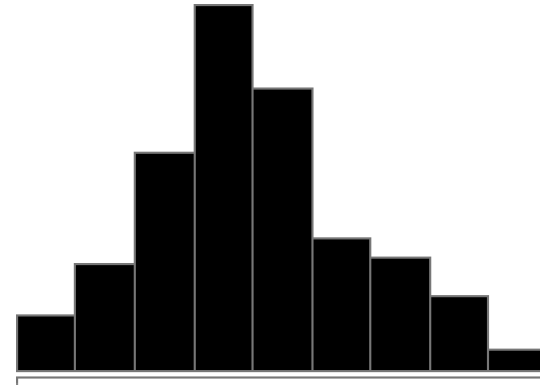
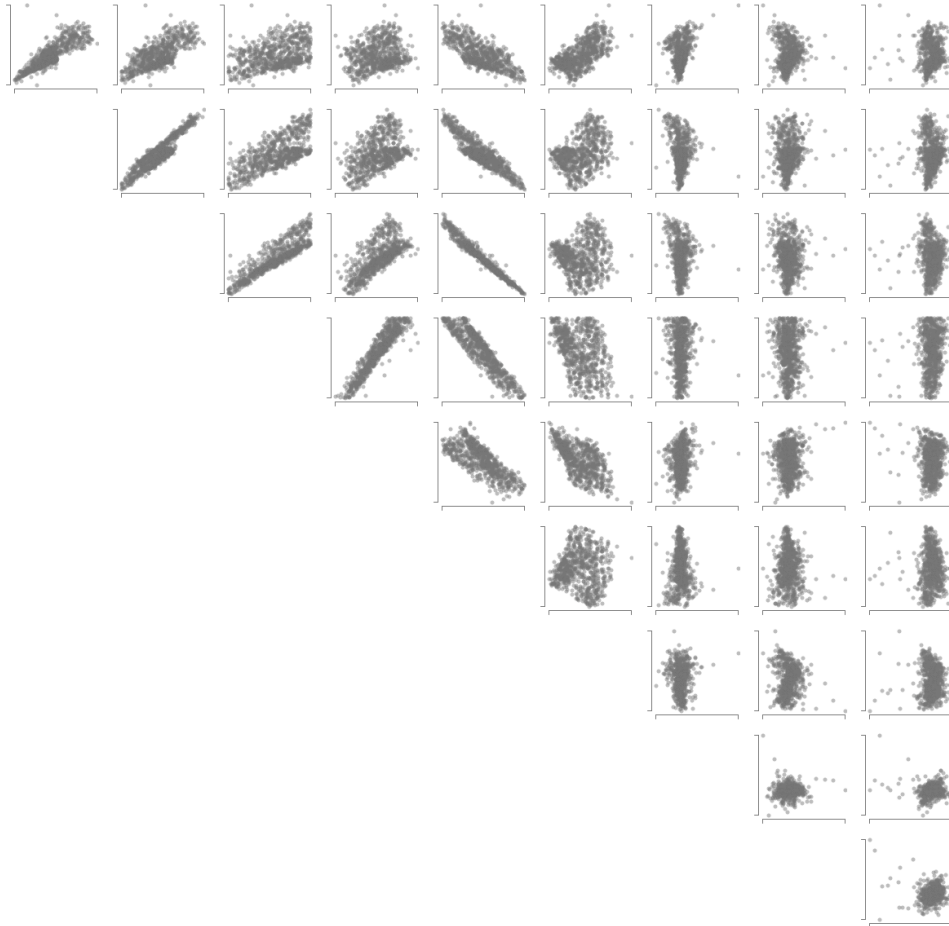
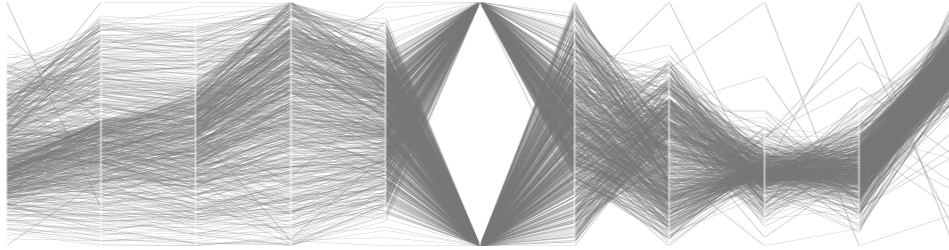
n uncorrelated variables

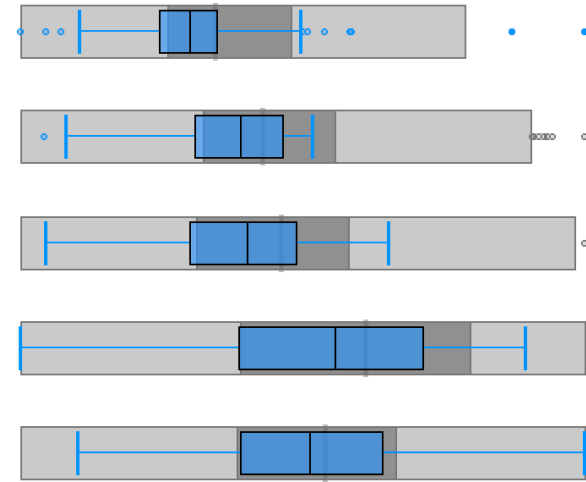
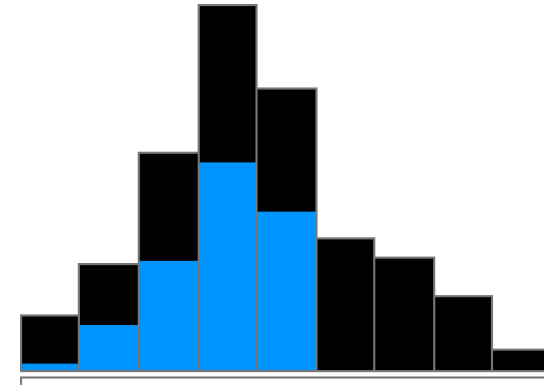
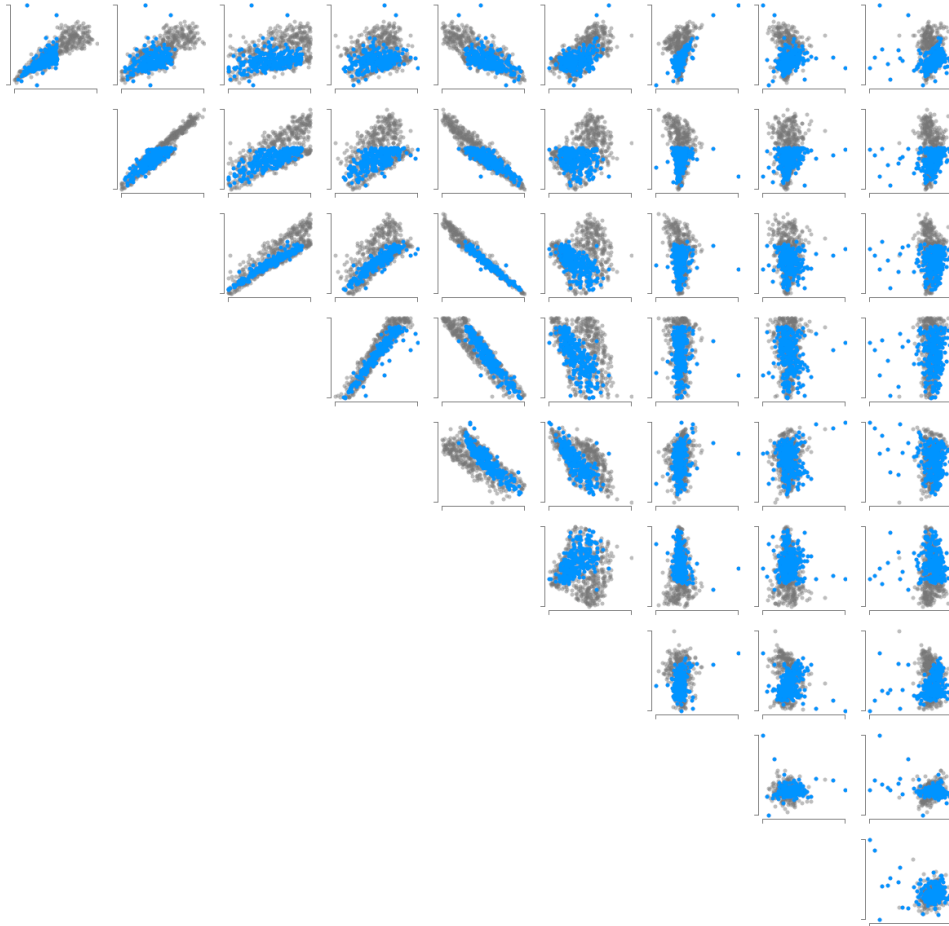
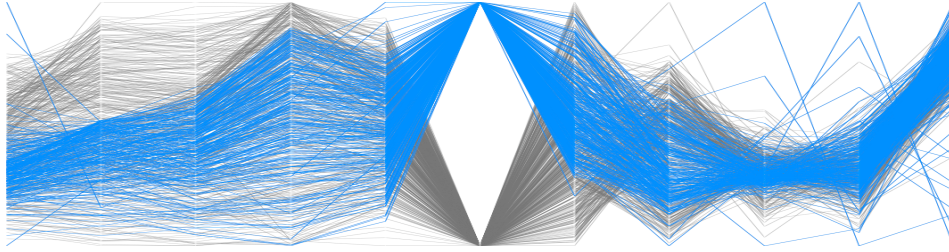
(akin decomposition into orthogonal element basis)



$\{ \rightarrow \text{all together} \}$







dashboards

information summaries

lead to suggested actions

exploratory environments

mondrian (free: R* back-end)

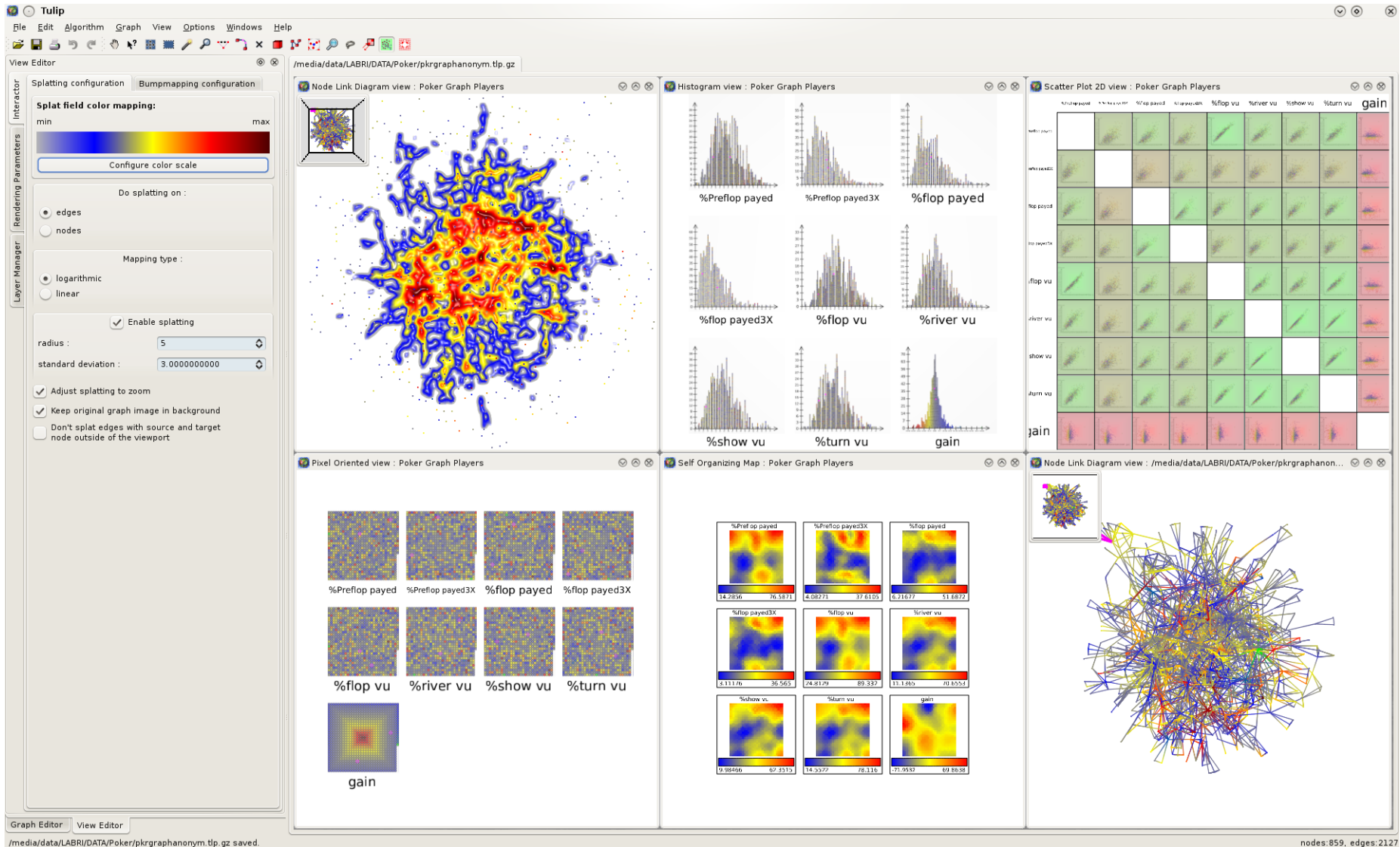
topcat (free: Virtual Observatory tools)

tableau (corporate, from stanford's data cubes+polaris)

tulip (relational data)

“visual analytics” driven by department of homeland security

(see past VAST contest)



dashboards

information summaries

lead to suggested actions

exploratory environments

mondrian (free: R* back-end)

topcat (free: Virtual Observatory tools)

tableau (corporate, from stanford's data cubes+polaris)

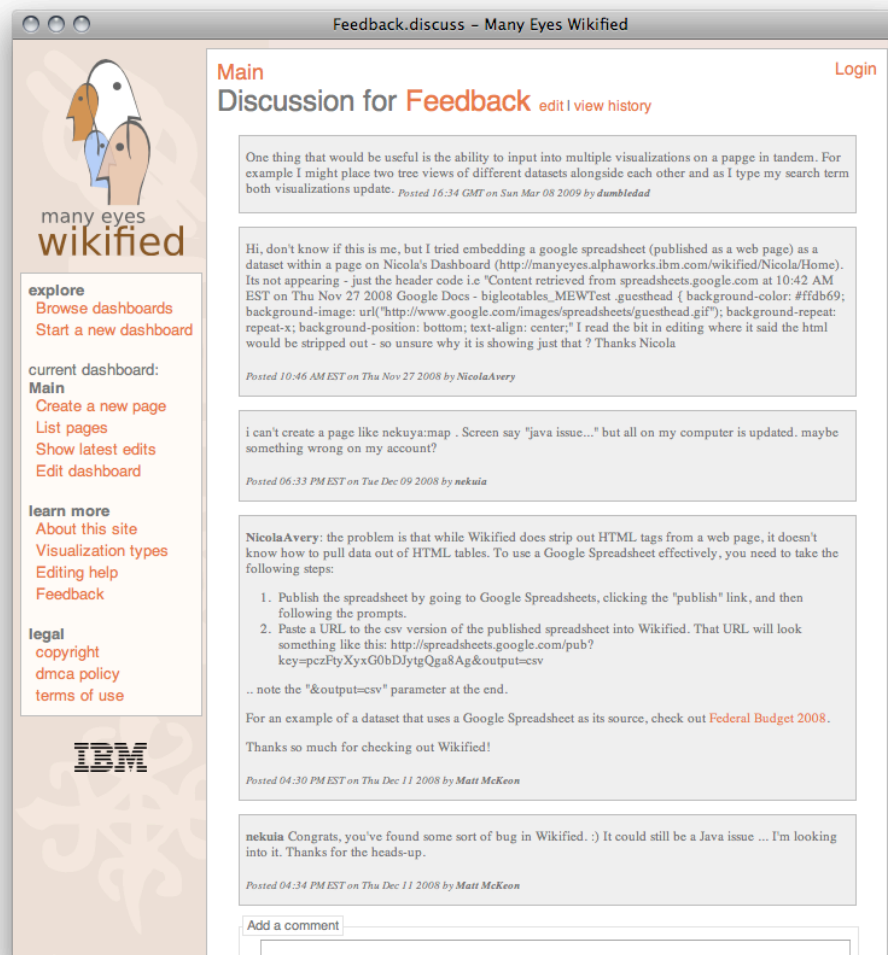
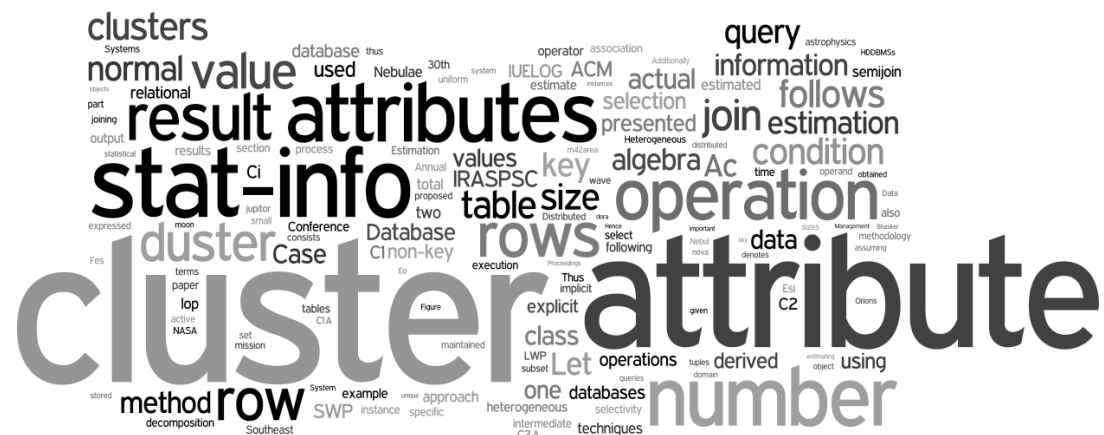
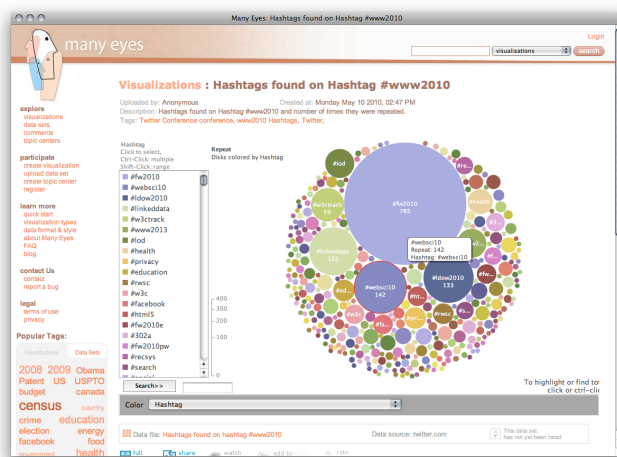
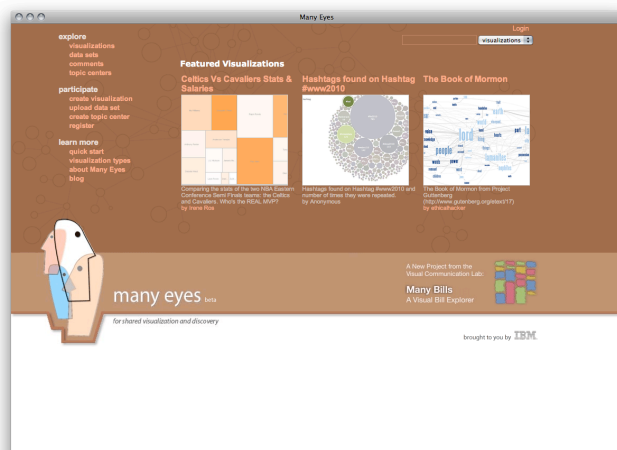
tulip (relational data)

“visual analytics” driven by department of homeland security
(see past VAST contest)

+ WWW

ibm's dashiki (data management system) + *many eyes* (wiki frontend)

(stanford's *protoviz* javascript+svg api)





many eyes
wikified

explore

[Browse dashboards](#)
[Start a new dashboard](#)

current dashboard: Learning Designer Results 30

[Create a new page](#)
[List pages](#)
[Show latest edits](#)
[Edit dashboard](#)

learn more

[About this site](#)
[Visualization types](#)
[Editing help](#)
[Feedback](#)

legal

[copyright](#)
[dmca policy](#)
[terms of use](#)

IBM

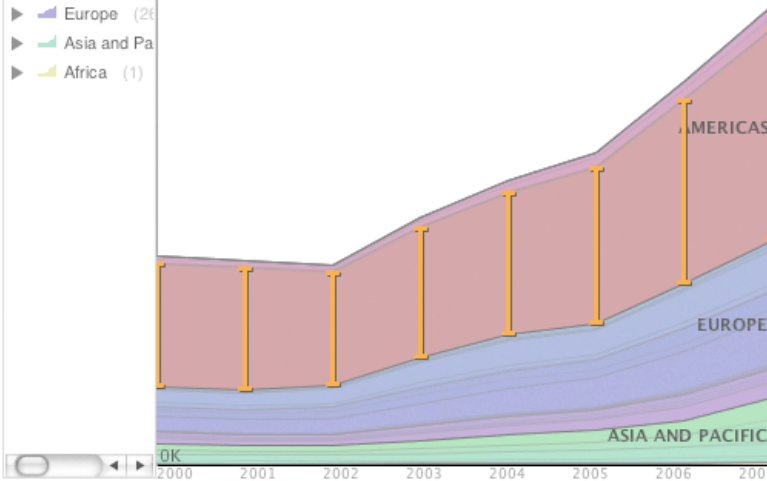
Main

Mutual Fund Holdings Worldwide [edit](#) | [discuss](#) | [share](#) | [view history](#)

Region Country

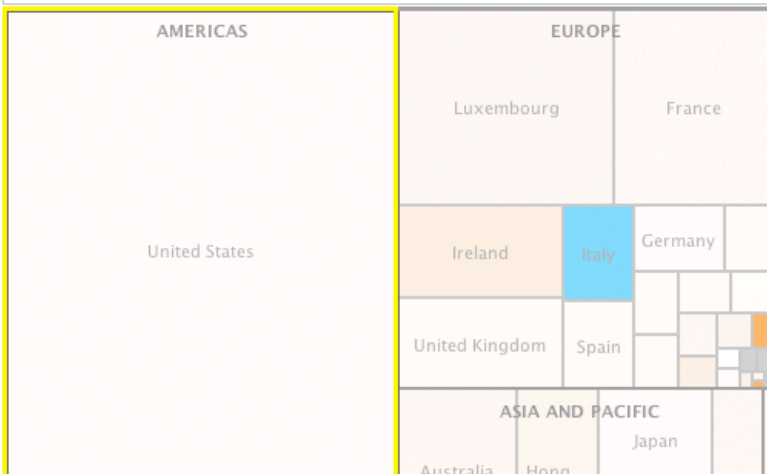
All (44)
Americas (26)
Europe (26)
Asia and Pa (1)
Africa (1)

30,000K



Mutual Fund Holdings:Chart | [edit](#) | [view data](#)

Region Country



Loans per 1000 housing units



[SubprimeData:map](#) | [edit](#) | [view data](#) Map data © 2007 ESRI

Main

History for SubprimeData:map

When	Edited by	Comment	Thumbnail
09:35 PM EDT on Fri Mar 27 2009	Administrator		
09:34 PM EDT on Fri Mar 27 2009	Administrator		
09:33 PM EDT on Fri Mar 27 2009	Administrator		
01:10 PM EDT on Thu Oct 16 2008	Administrator		

data analysis

interactive + well designed visuals

r ✚ dashiki ✚ facebook ✚ ipad ✚ skype

intuitive interaction

collaboration

notes

best answers ➡ task dependent analytics

next best: mindful visualization ➡ understanding

possible: n variables ➡ single plot

note: mind basic design principles

further extending for highly complex data exploration:

cross referenced views + interativity

note: many tools + techniques out there

follow the data

(complement+explore+exploit ➡ customize)

questions?

corrections:

notes on visualization
santiago v lombeyda

visual exploration  *answer questions*
(even if questions are not clearly defined?)

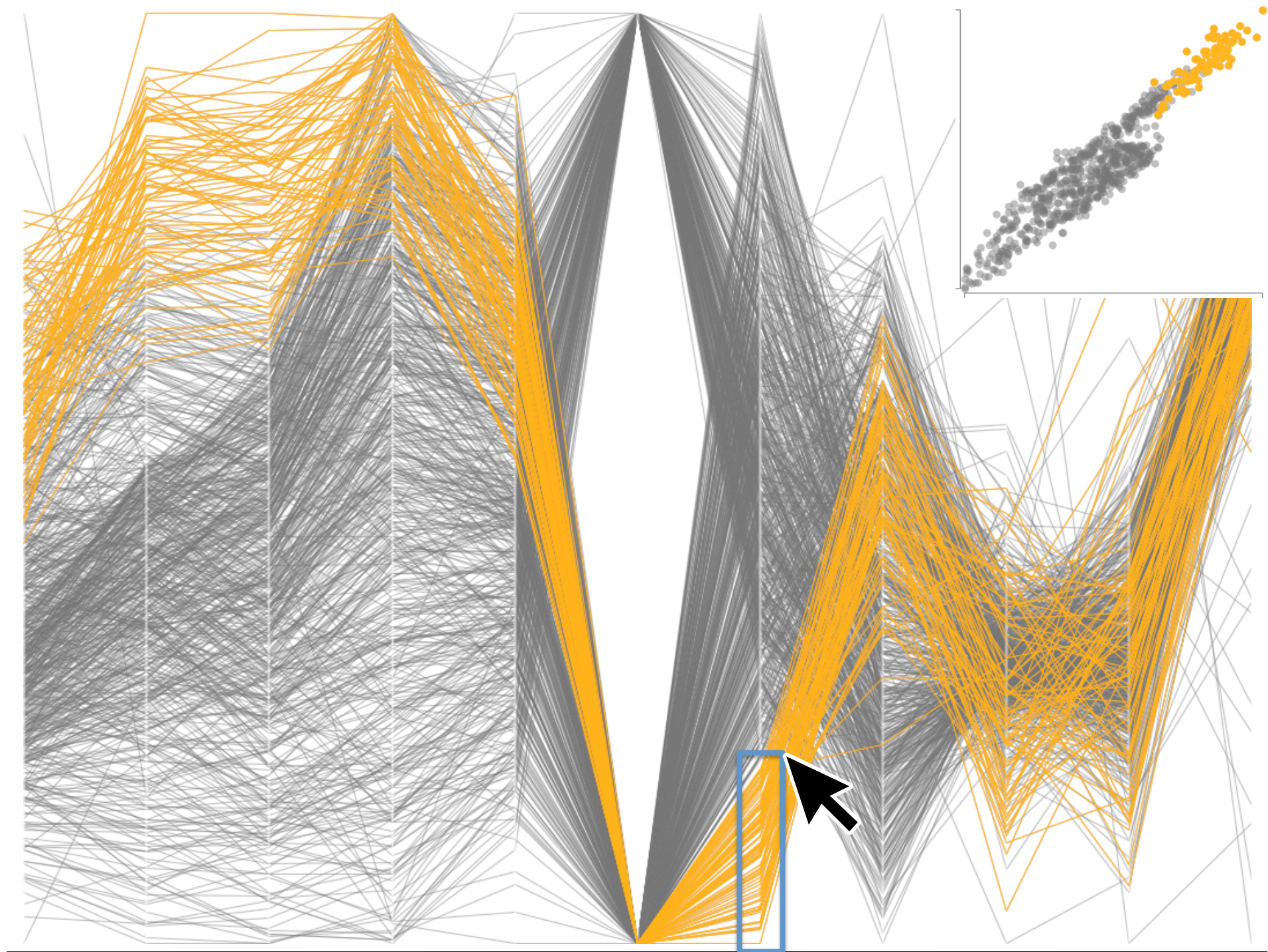
 tools address specific task/question

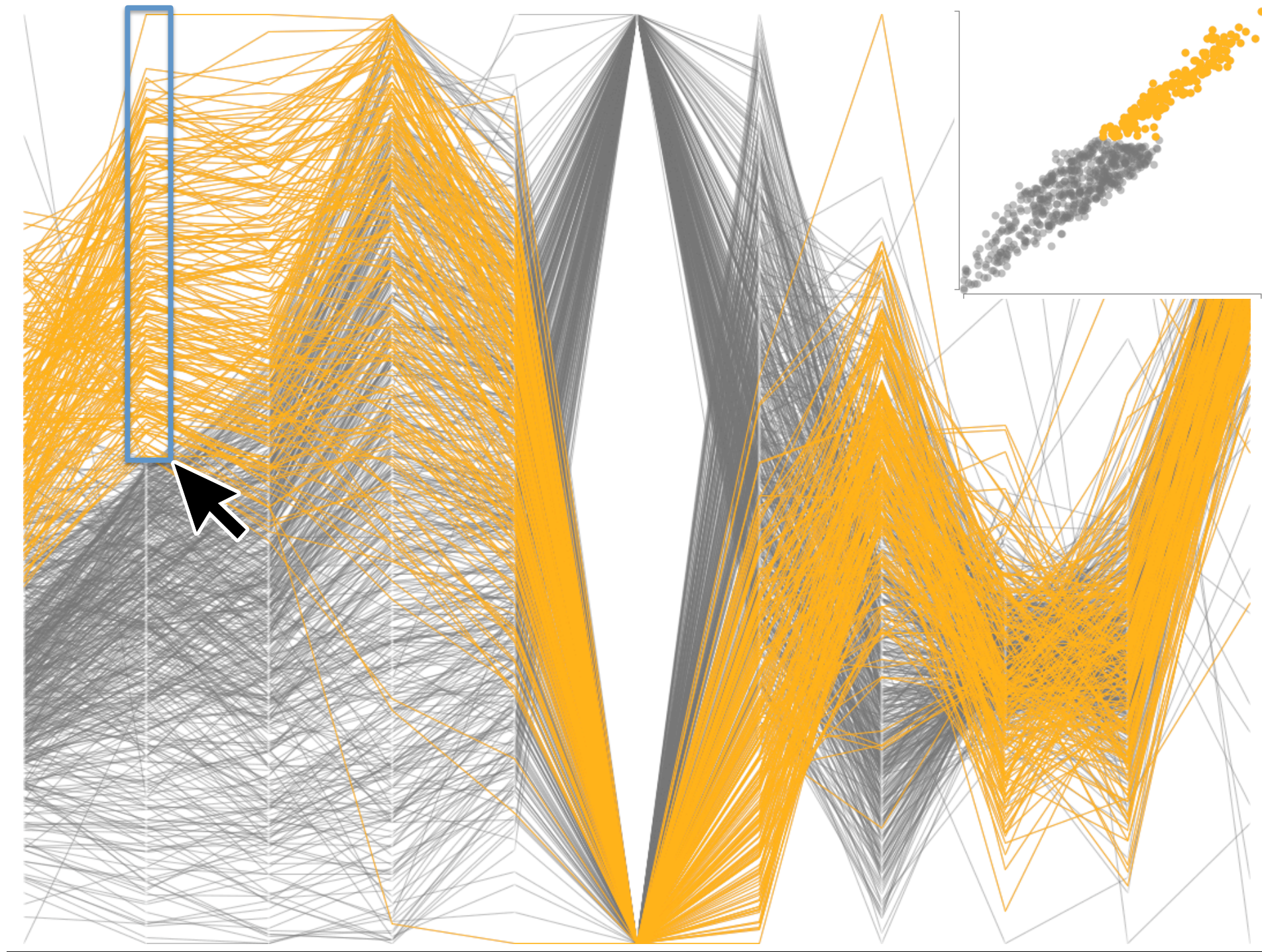
(visualization is bound by tools... right?)

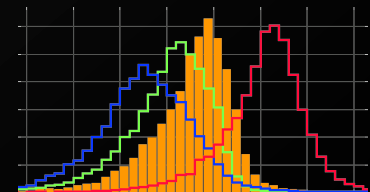
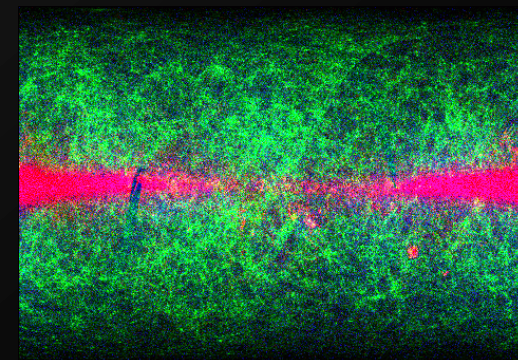
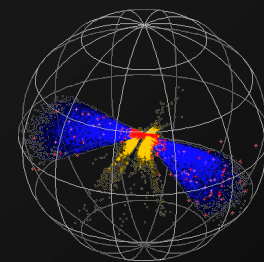
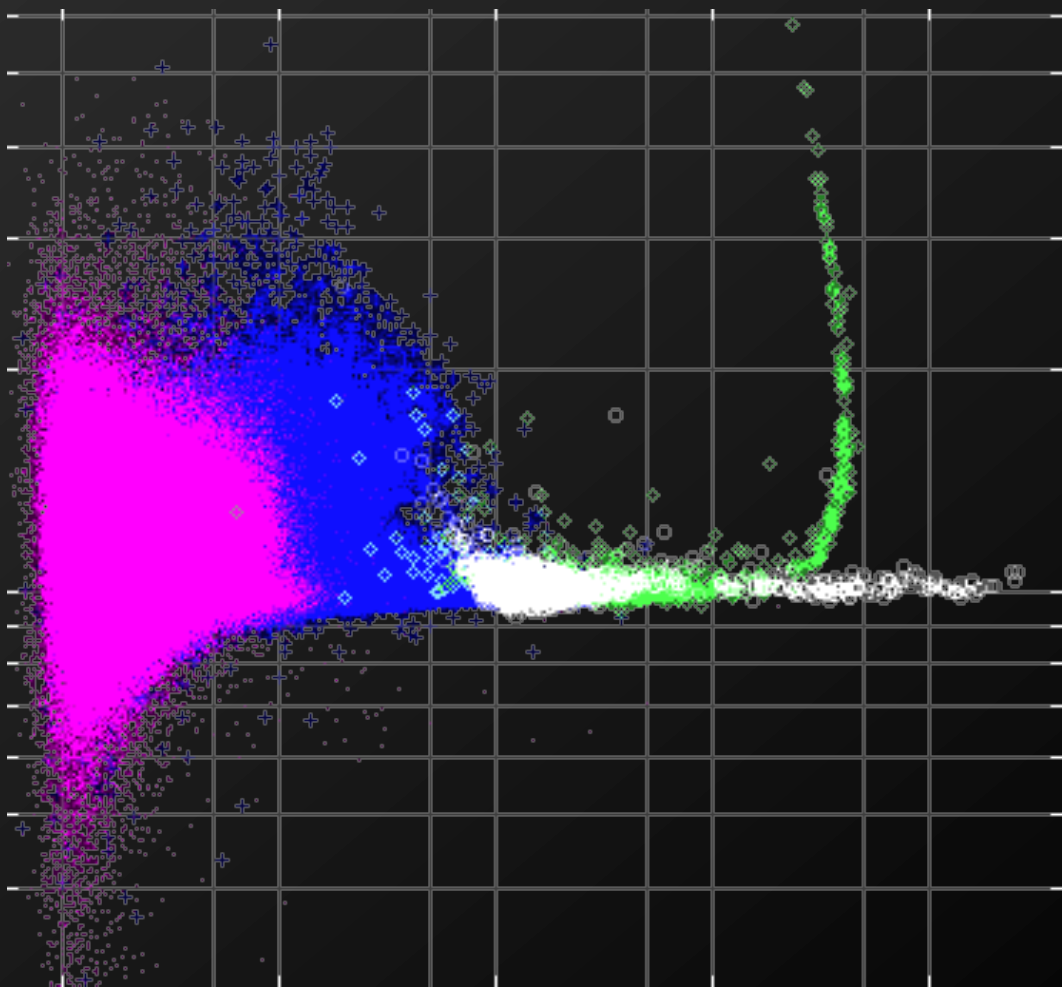
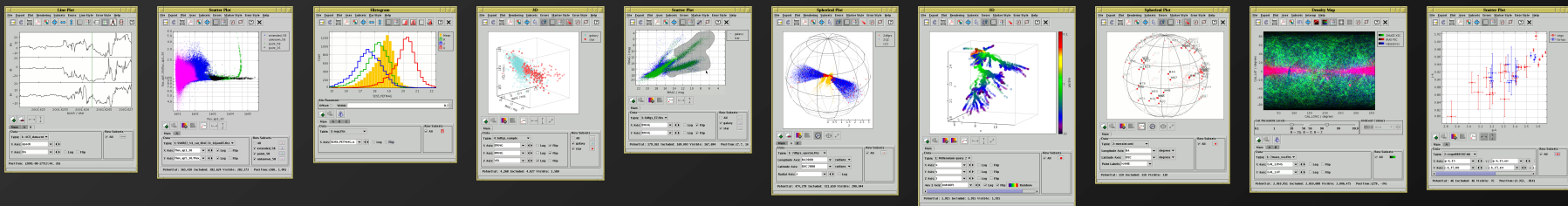
 cross referencing

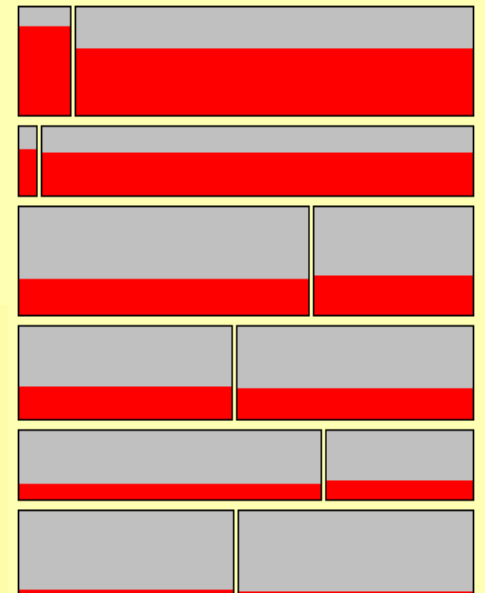
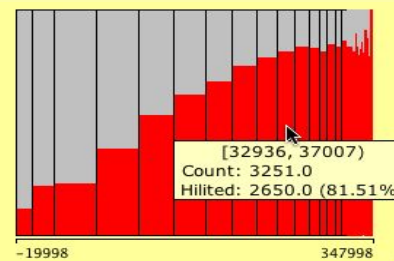
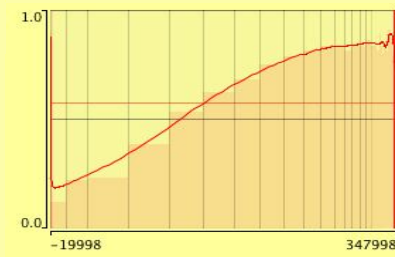
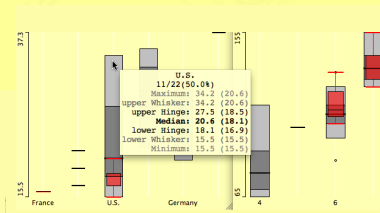
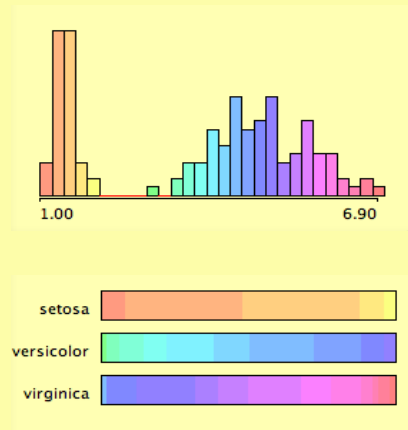
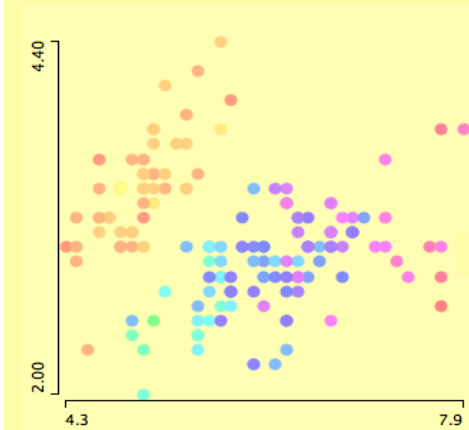
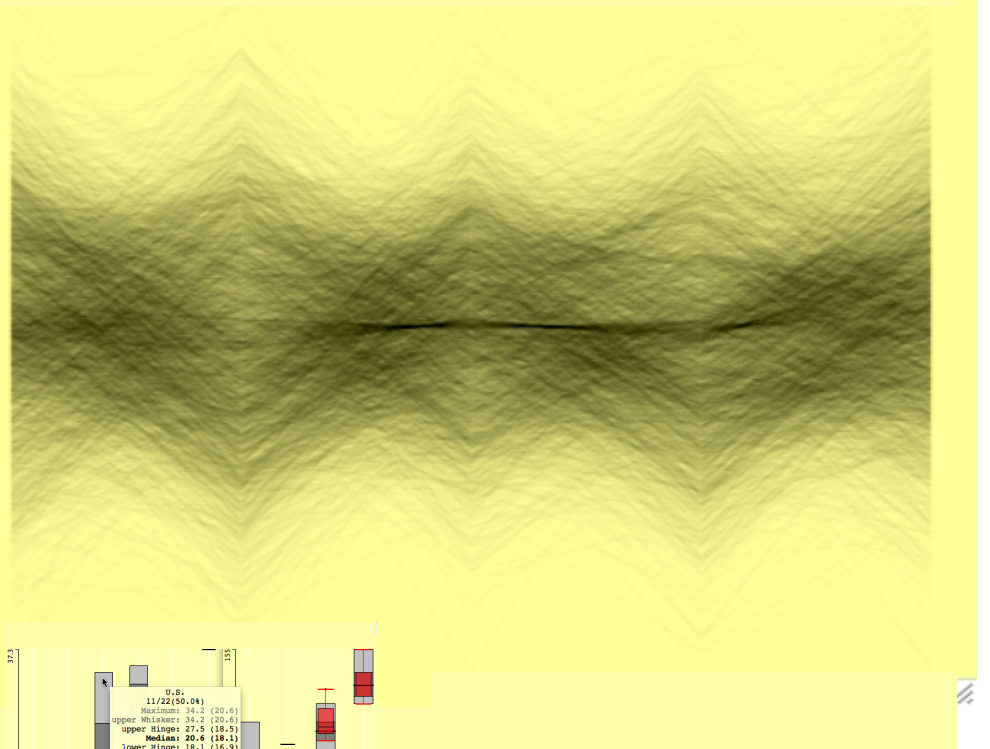
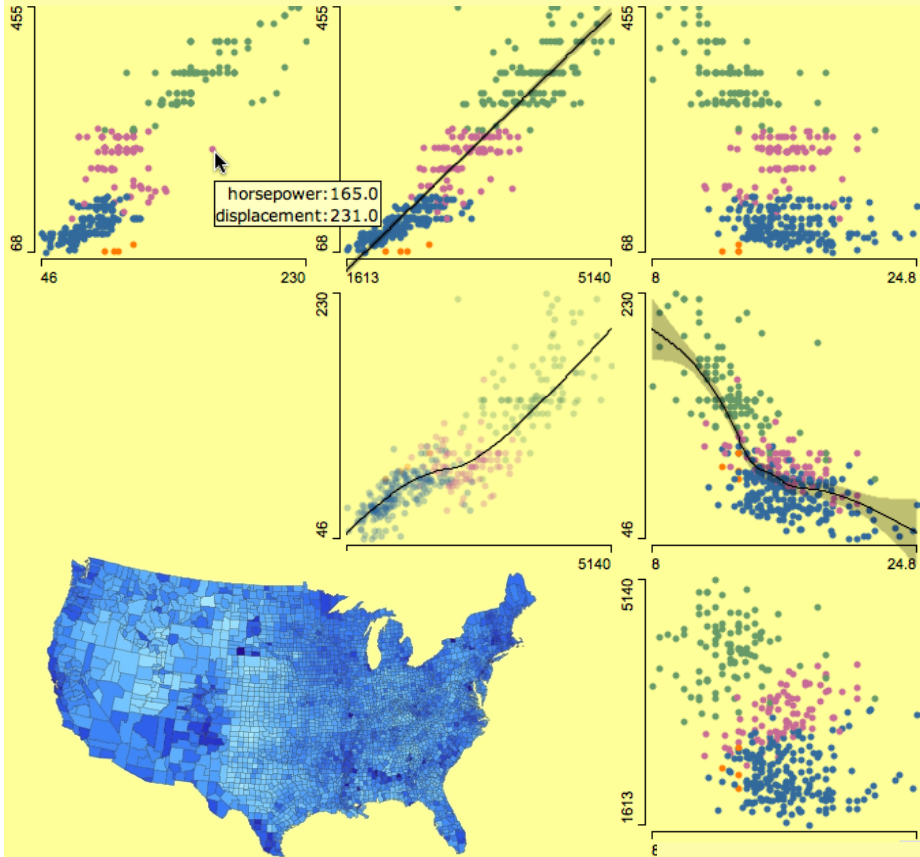
 interactivity

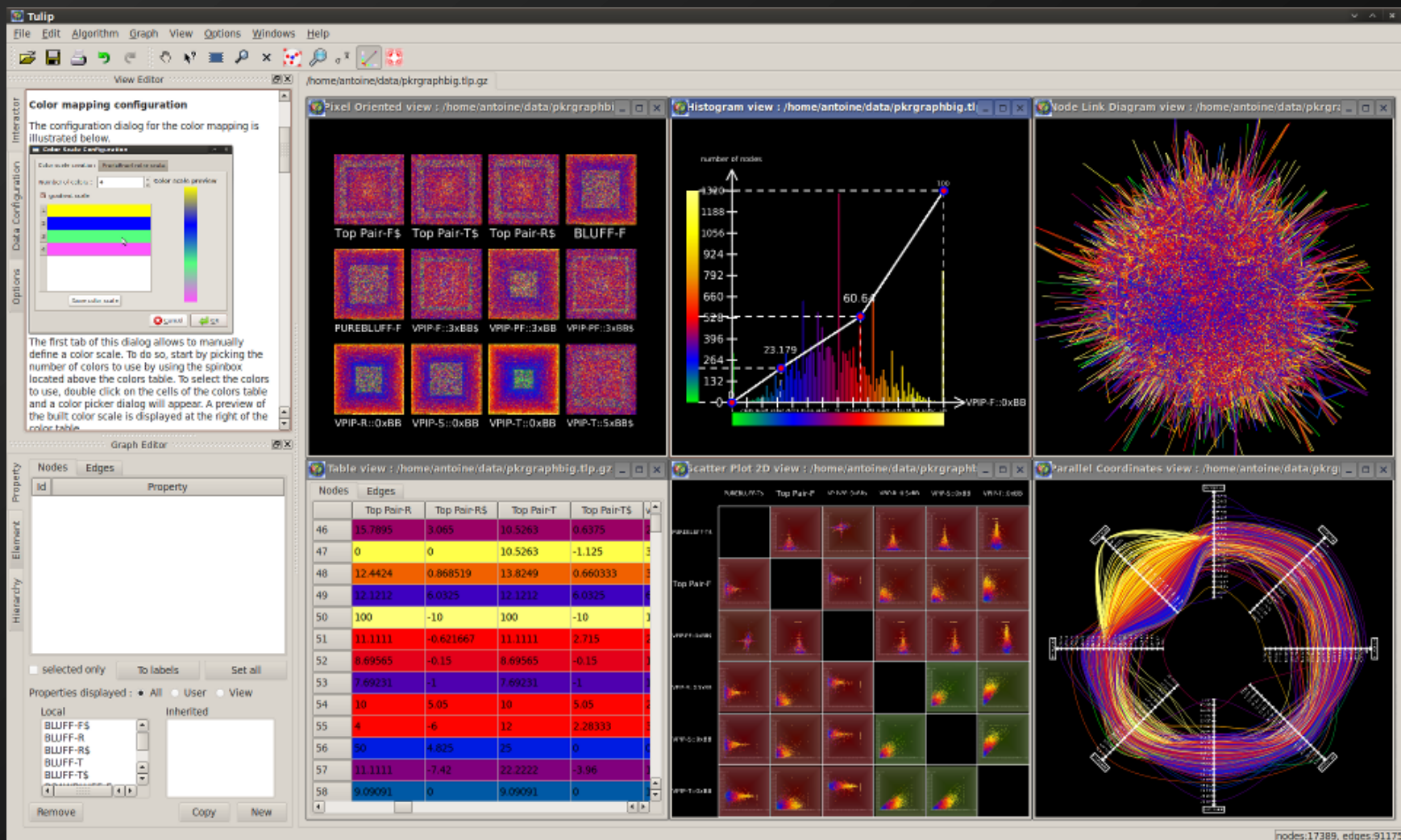
 *broader* set of questions addressable

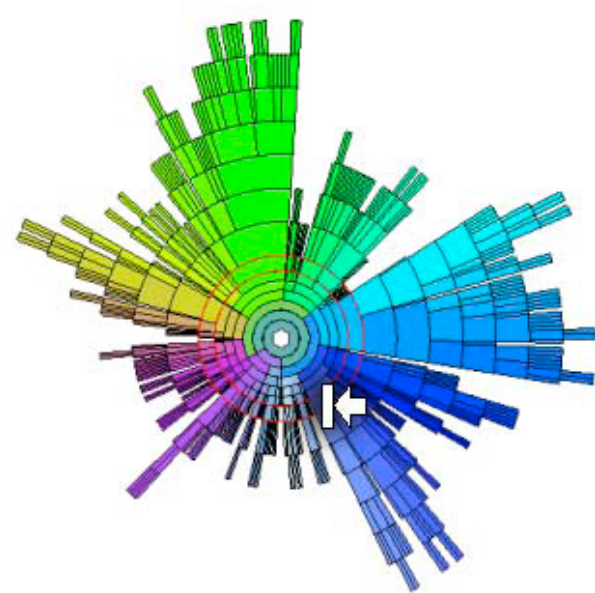
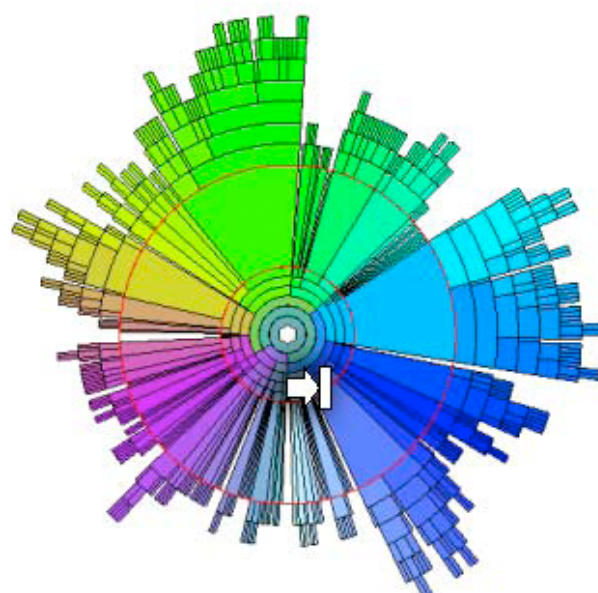
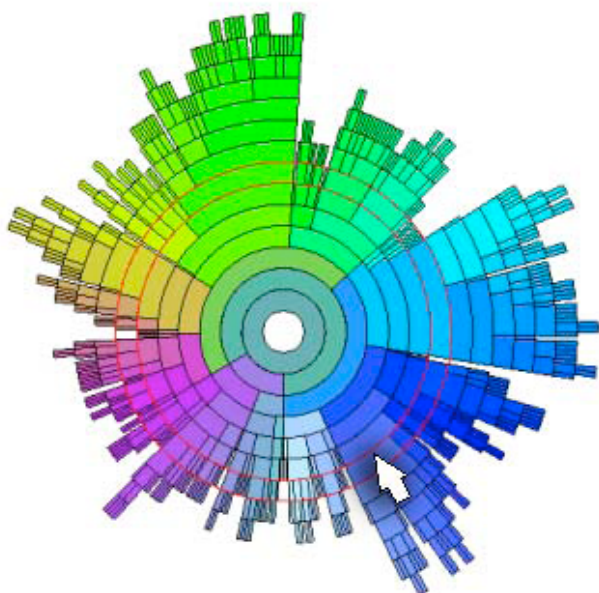
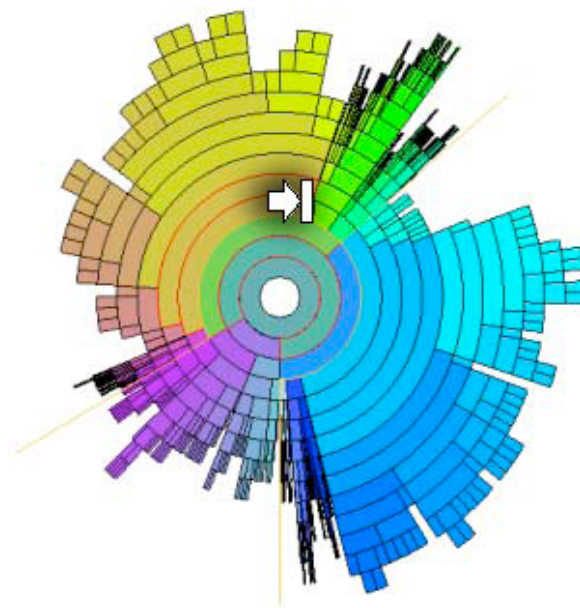
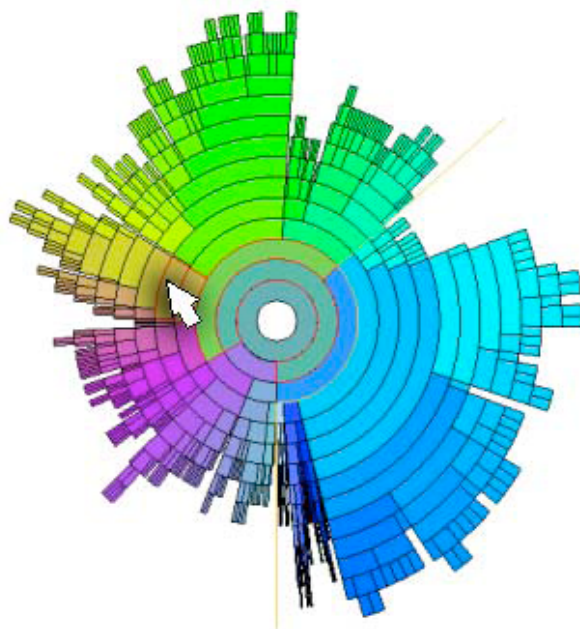
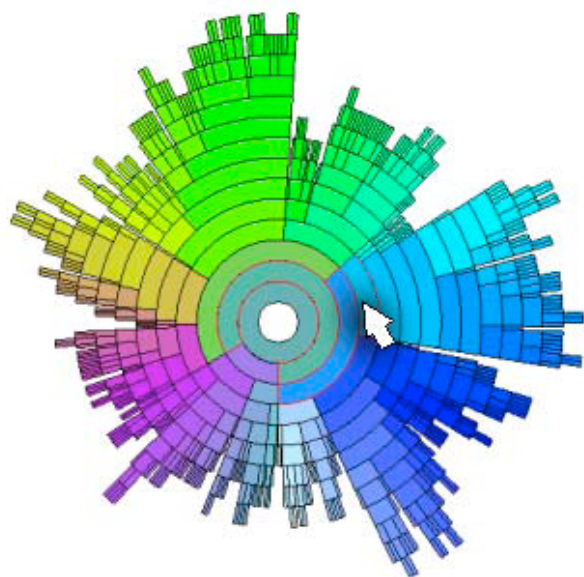












textpresso

text-mining system  scientific literature

(access to full text)

encapsulates: categories of biological concepts+classes

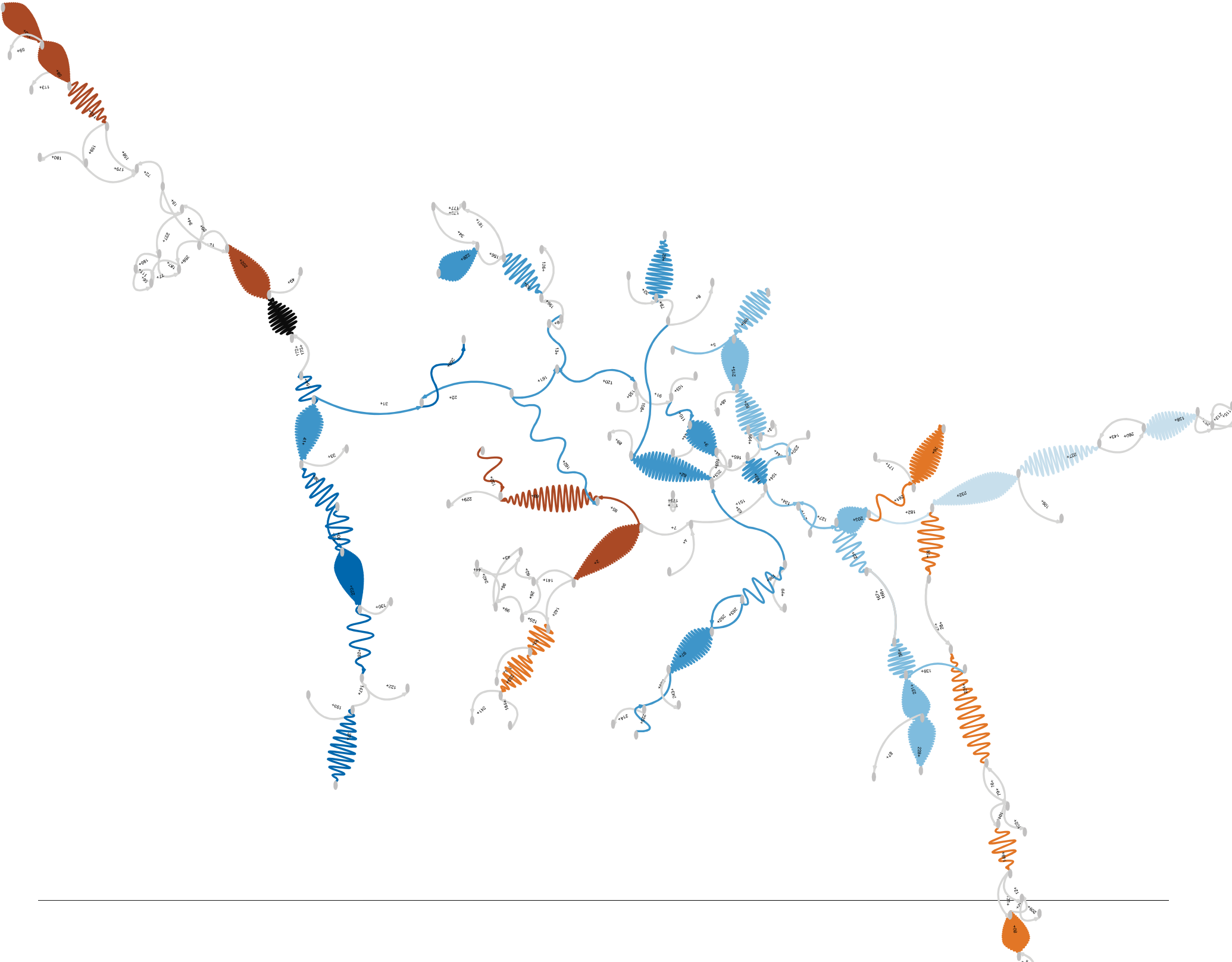
encodes: how two objects relate (association, regulation, ...)

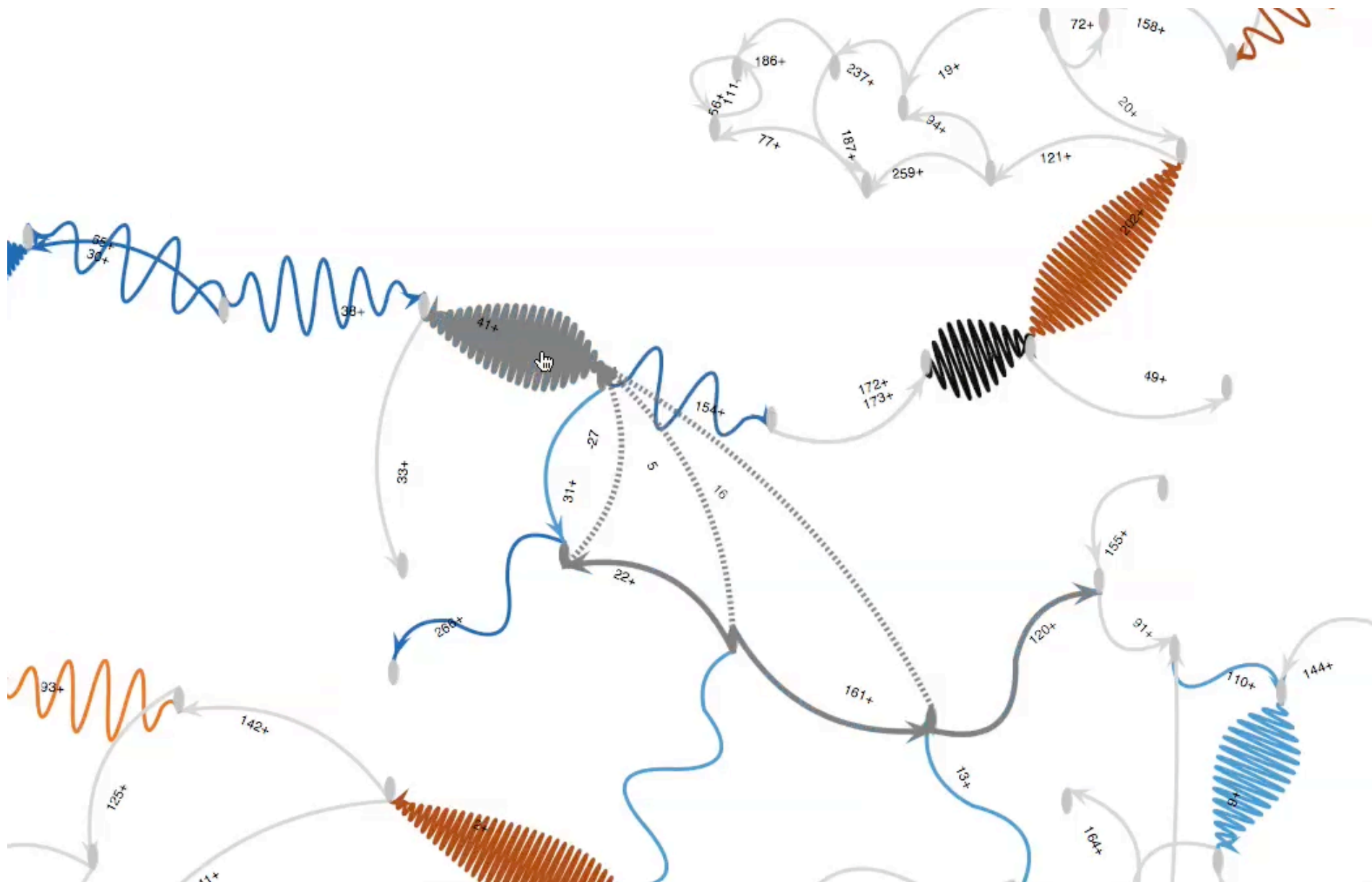
encodes: object description (methods, ...)

search engine

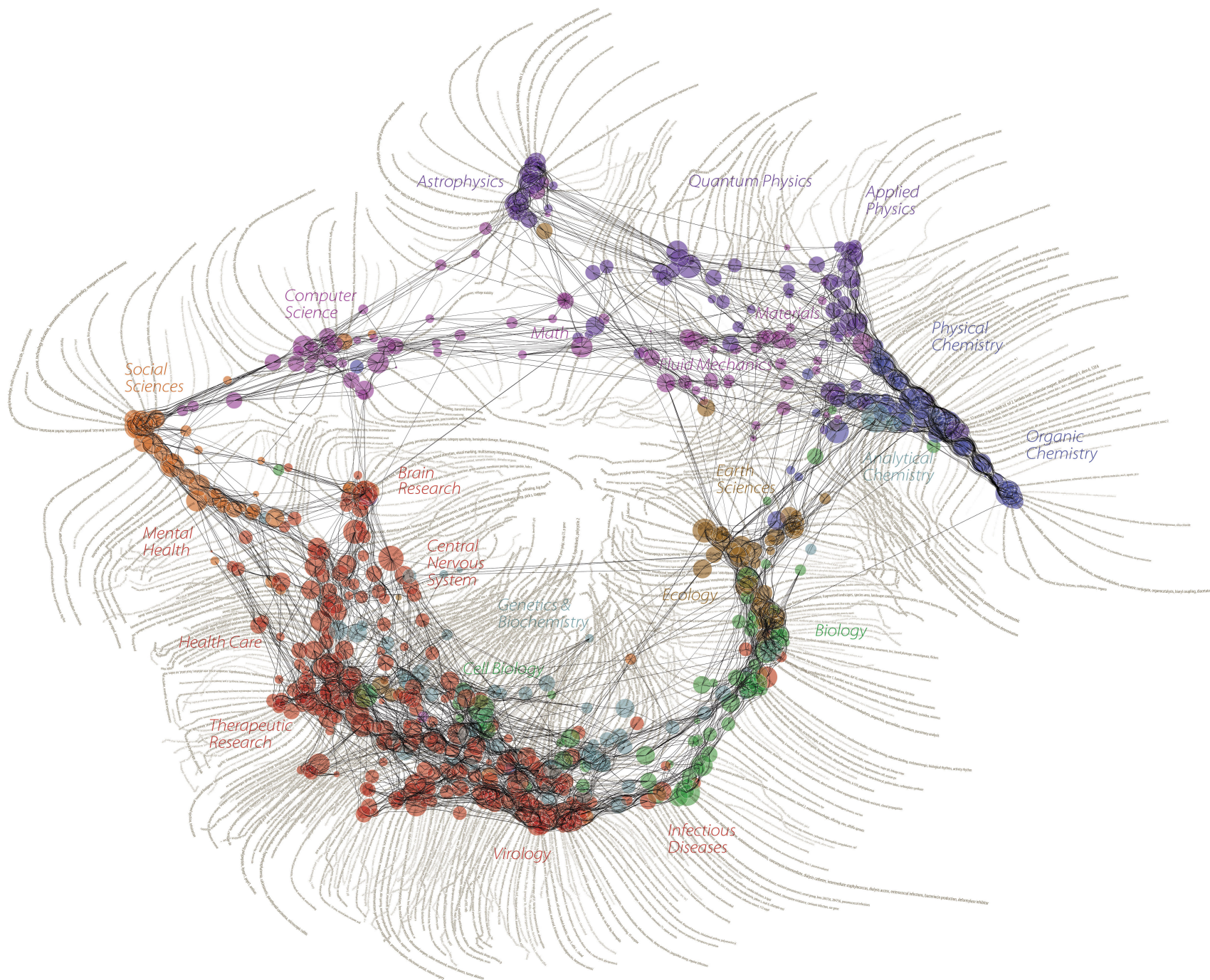
combination categories

keywords





Relationships Among Scientific Paradigms



[illegible][illegible][illegible]

