Connecting The Dots - Showing Relationships in Data and Beyond

Marc Streit[1], Hans-Jörg Schulz[2], Alexander Lex[3]

VisWeek Tutorial 2012

1. Johannes Kepler University Linz, Austria
2. University of Rostock, Germany
3. Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA

PART 1:
WHAT TO LINK?

Speaker: Hans-Jörg Schulz

Linking What Belongs Together

Fulfilling the criterion of being expressive, we want to link "stuff" that it is

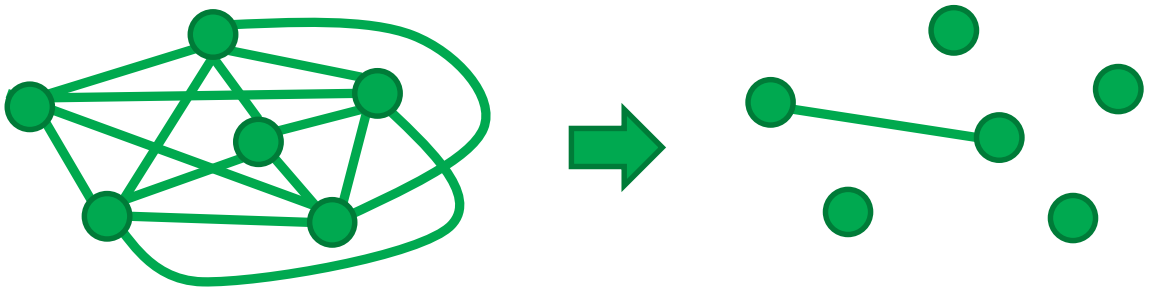
Related, Associated, Connected, Affiliated,…

"When two objects, qualities, classes, or attributes, viewed together by the mind, are seen under some connexion, that connexion is called a relation." —Augustus De Morgan (1858)

Linking What Belongs Together

Disclaimer: There appear to be cases in which the opposite is useful/done.

The EIRTEE*-Scenario



*Everything-Is-Related-To-Everything-Else

Relations

Definition:

Relations assign true/false to a *k*-tuple.

Aspects to consider:

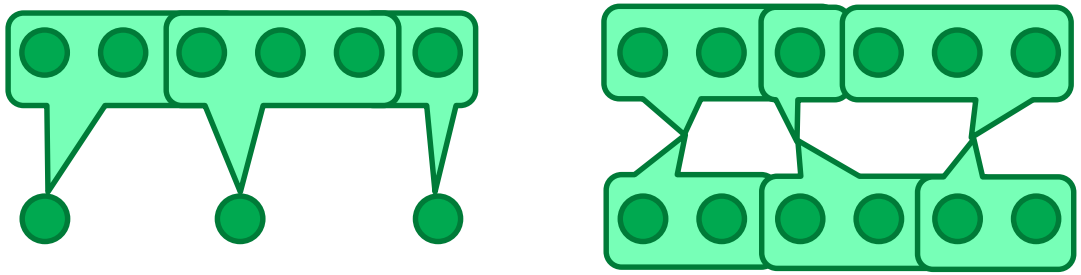
- the cardinality *k*
- the elements of the tuple
- the domain in which the relation is defined

Relations: Cardinality

Cases:

- *k=0*: tautology (TRUE) + contradiction (FALSE)
- *k=1*: unary (property)
- *k=2*: binary (is not necessarily a 1:1 relation!)

- *k>2*: n-ary with n>2



Relations: Elements

We define relations over (some part of) data.

We assume a simple hierarchical data model:

- $\{attrib_1,\ldots,attrib_i\}$ = data item
- $\{item_1,\ldots,item_j\}$ = data cluster
- $\{cluster_1,\ldots,cluster_k\}$ = data set
- $\{set_1,\ldots,set_l\}$ = data landscape

The tuple elements are drawn from these levels.

Relations: Elements

A set of elements is defined by a granularity and a scope.

While the granularity defines the level of detail of the relation, scope defines its extent.

For example, relations can be established between data clusters (granularity), which are drawn either from an individual data set or from multiple data sets of a data landscape (scope).

Relation Scope level > Relation Granularity

Relations: Domain

While the relation is defined over the data domain, it may stem from any of the following:

- Data Domain
- View Domain
- Interaction Domain

To identify the original Relationship Domain can sometimes be hard: Is data related because it is jointly interacted with, or is it jointly interacted with because it is related?

Relations put Together

- Cardinality
  - binary
  - n-ary with n>2
- Domain
  - Data
  - View
  - Interaction
- Elements = Granularity + Scope
  - Data attributes
  - Data items
  - Data clusters
  - Data sets
  - Data landscape

Relationships: Other Aspects

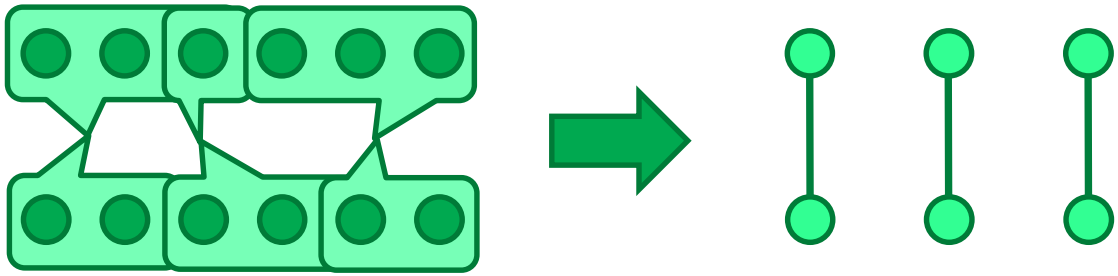
- Inherent vs. Derived
- Transitivity, Directionality, Strength
- Multiple relationships
  Example: coauthorship, citation, co-citation,…
- Running between different
  - Data tuples, Data tables, Data bases
  - Graphical objects, Views, Applications

Relationships: Interrelations

Under certain circumstances, it is possible to transform relationships. For example:

- n-ary relations on data item level
  → binary relations on data cluster level

(via the 1:n inclusion relation between items and cluster)



Relationships: Interrelations

Under certain circumstances, it is possible to transform relationships. For example:

- n-ary relations on data item level
  → binary relations on data cluster level

(via the 1:n inclusion relation between items and cluster)

- relations derived from interaction domain
  → relations from data domain, if interaction logs are considered as additional data set

Relationships: Examples

Example #	Description	Domain	Granularity	Scope	Cardinality
1	ARGOIs	Data	Attributes	Data set	binary
2	Graphs	Data	Items	Data set	binary
3	Hypergraphs	Data	Items	Data set	n-ary
4	Clustered Data	Data	Items	Data set	n-ary
5	Matchmaker /VisBricks	Data	Clusters	Data set	binary
6	StratomeX	Data	Clusters	Landscape	binary
7	StratomeX: DVI	Data	Data Sets	Landscape	binary
8	Spatial Treemaps	View	Attributes	Data set	binary
9	Stack‘n‘Flip	Interaction	Data sets	Landscape	binary

Example #1: ARGOIs

(Domain: Data, Elements: Attributes in Data Set, Cardinality: binary)

- Relationship: two attributes are related, iff they belong to the same data tuple
- Common visual representation: ParCoords

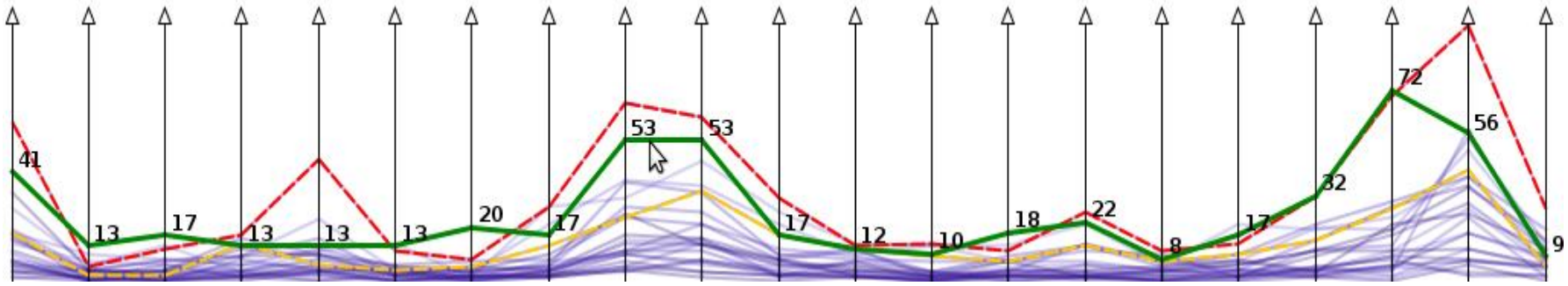
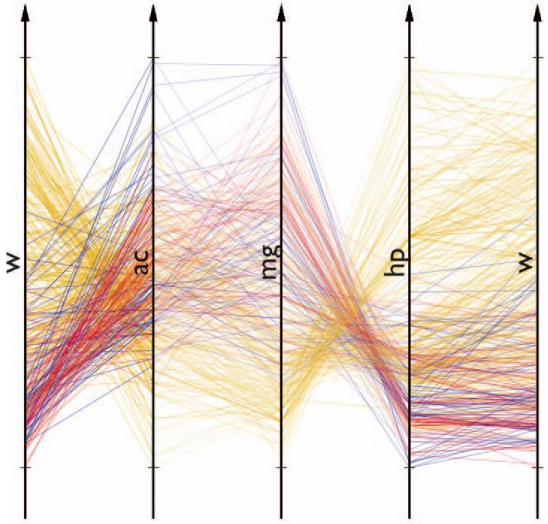


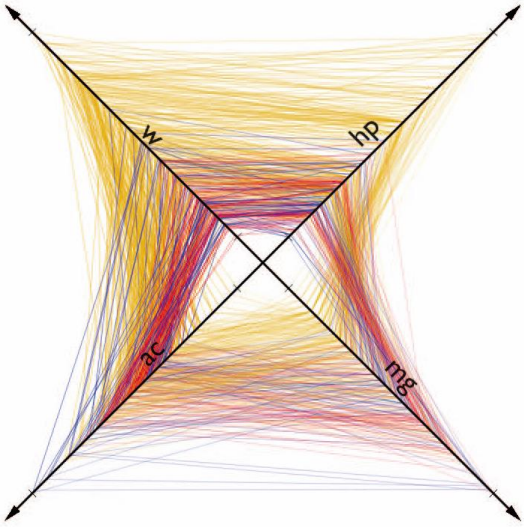
Image taken from Tominski+Schulz (2012)

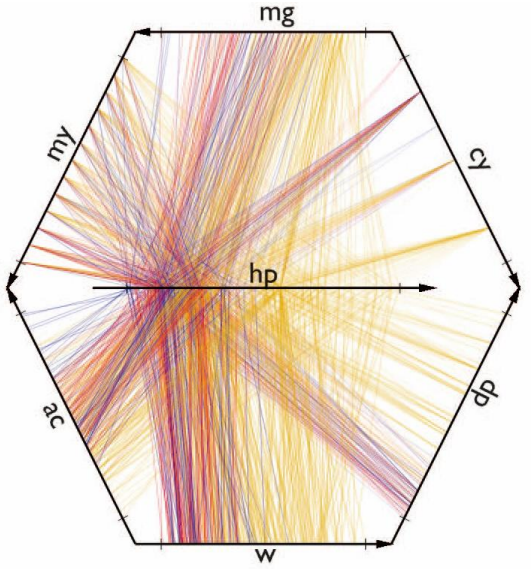
Example #1: ARGOIs

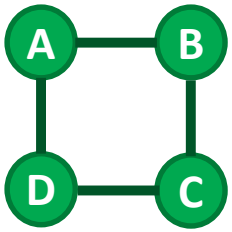
(Domain: Data, Elements: Attributes in Data Set, Cardinality: binary)

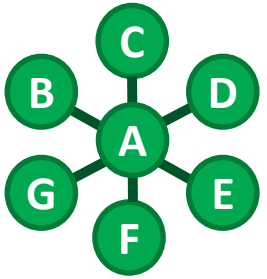
Generalization by Claessen+Wijk (2011):









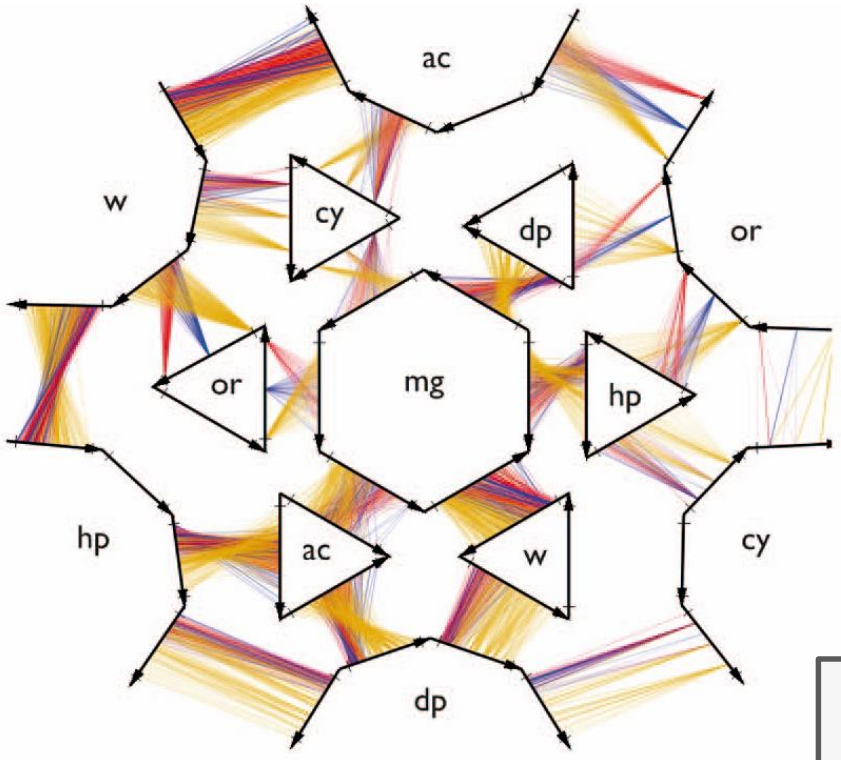


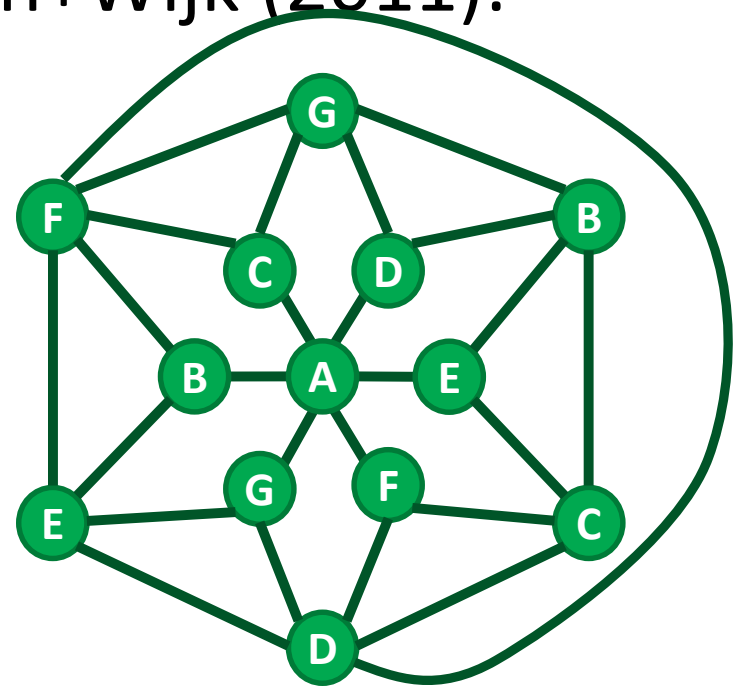
Attribute Relation Graphs of Interest ARGOIs

Example #1: ARGOIs

(Domain: Data, Elements: Attributes in Data Set, Cardinality: binary)

Generalization by Claessen+Wijk (2011):





Straight-line planarity required!

Example #2: Graphs

(Domain: Data, Elements: Items in Data Set, Cardinality: binary)

- Relationship: two items are related, iff there exists an edge between them
- Common visual representation: Node-Link-Diagram

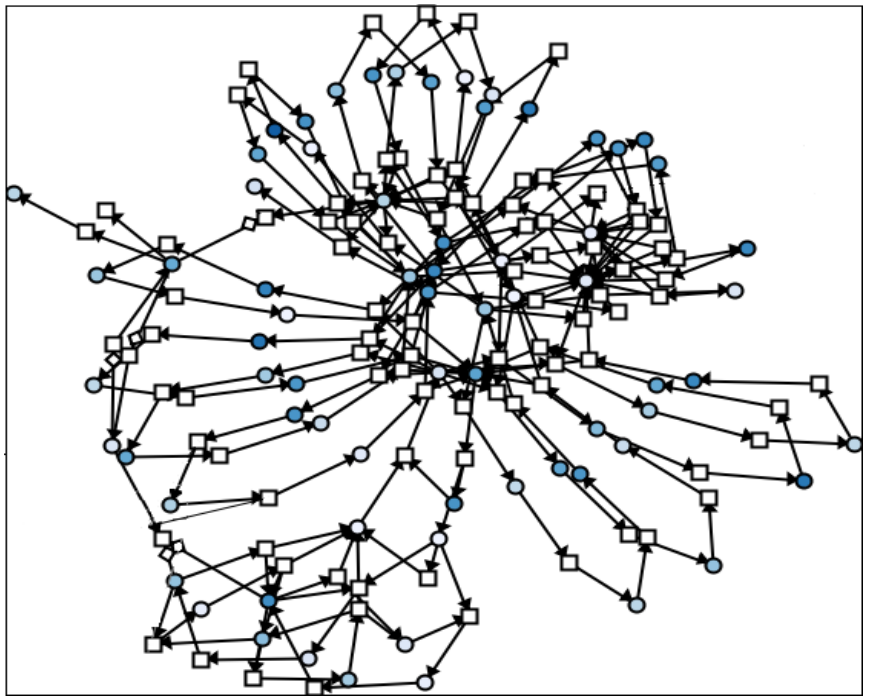


Image taken from
Hadlak, Schulz, Schumann (2011)

Example #3: Hypergraphs

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- Relationship: a number of items are related, iff there exists a hyperedge between them
- Common visual representation: Euler-Diagram

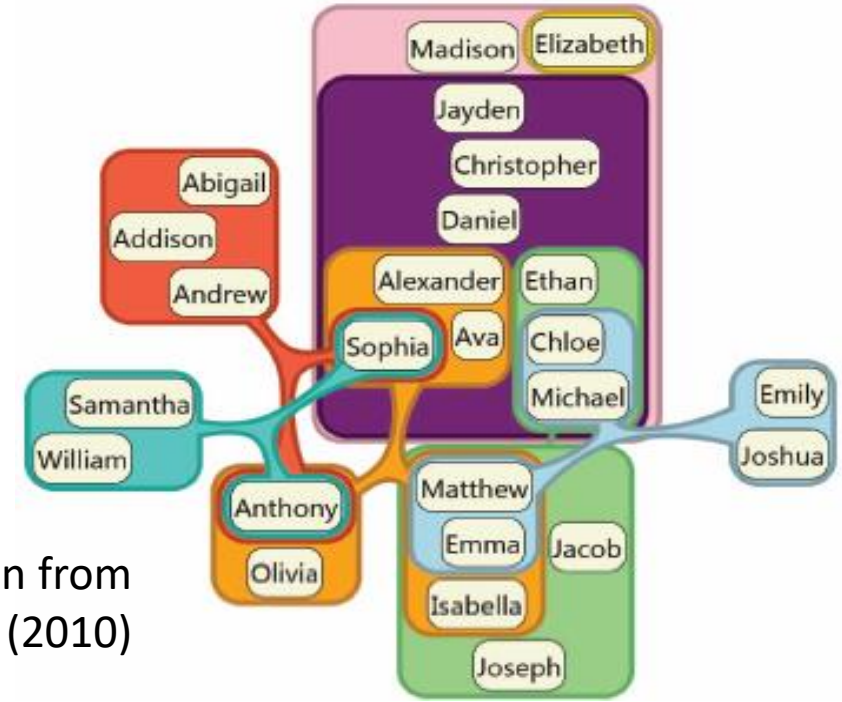
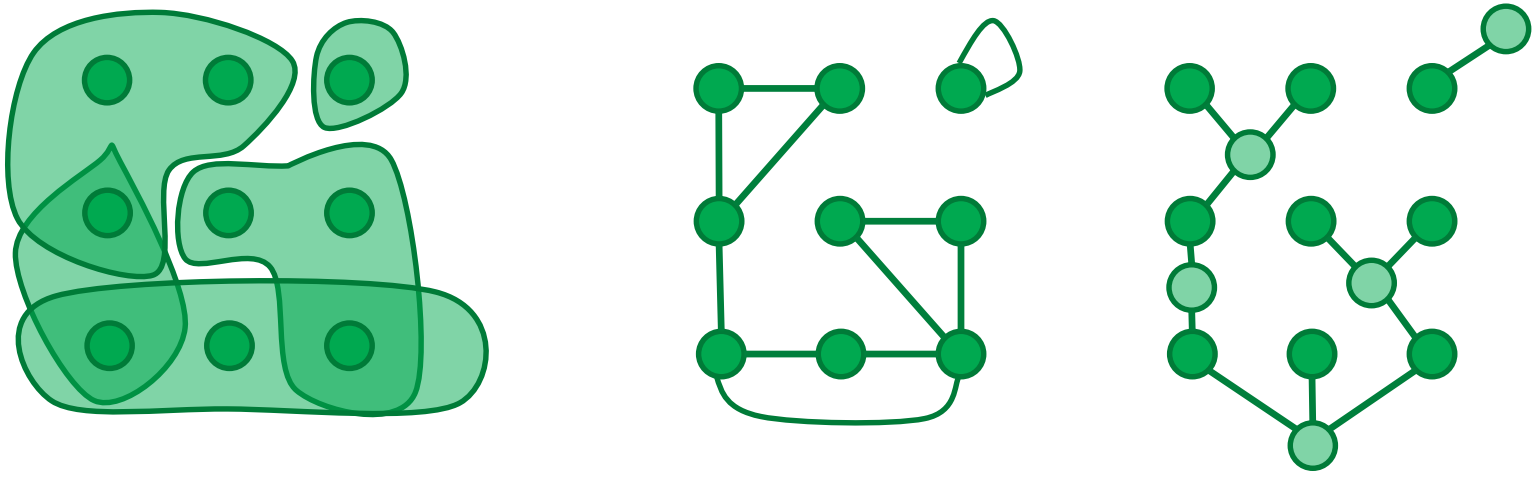


Image taken from Riche+Dwyer (2010)

Example #3: Hypergraphs

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- Recall interrelationships
- Hypergraphs can be transformed into regular graphs



Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- Relationship: a number of items are related, iff they belong to the same cluster
- Common visual representation: Scatterplot

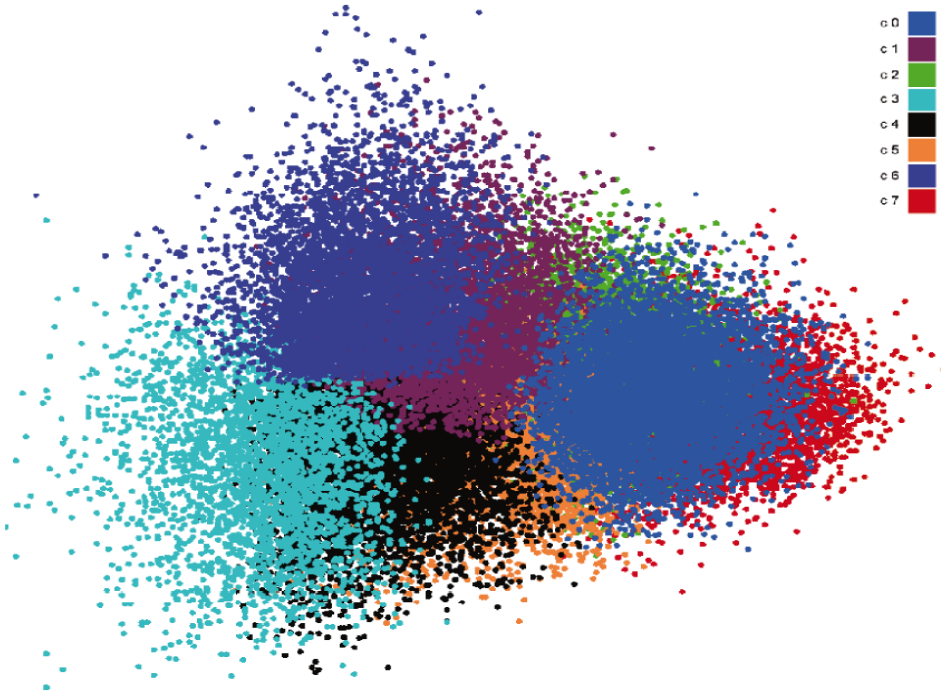


Image taken from Luboschik et al. (2010)

Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- *Clustering is the division of data into groups of similar objects.*
- **given:** a (dis-)similarity measure/matrix
  - n-dimensional, numerical data: Euclidean Distance
  - network data: Graph-theoretic Distance
  - strings of text: Edit Distance
- **sought:** a grouping of the data w.r.t. that measure

Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- What makes a good grouping?
  - **Compact:** elements in cluster are similar
  - **Separated:** clusters are different
  - **Balanced:** cluster membership is equally probable
  - **Parsimonious:** much fewer clusters than data objects

Source: Cosma Shalizi (2009)

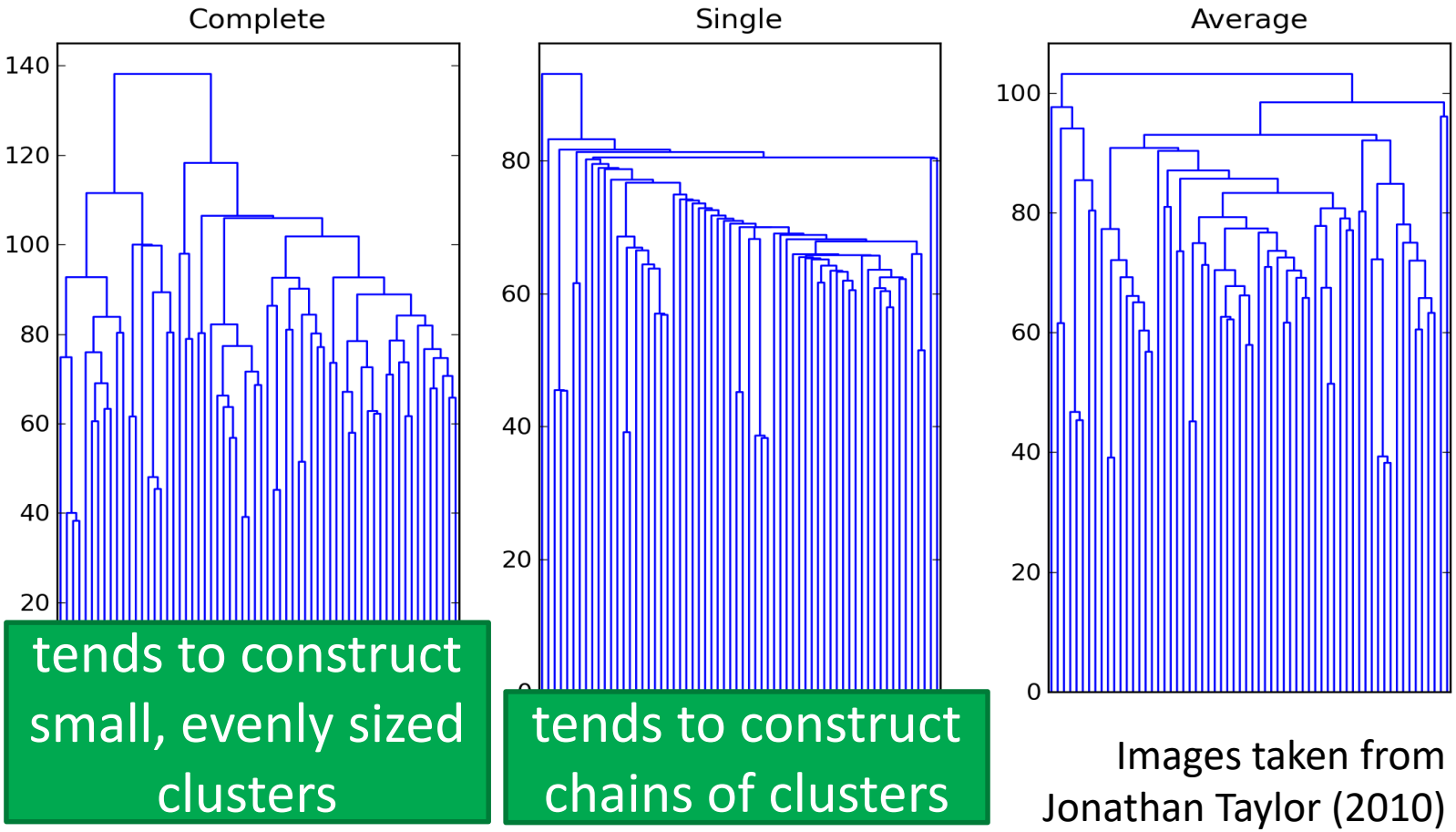
Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- Directionality of the clustering:
  - **Top-down:** divisive
  - **Bottom-up:** agglomerative
- Linkage metrics:
  - **Single Linkage:** nearest neighbor
  - **Complete Linkage:** farthest neighbor
  - **Average Linkage:** all neighbors

Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)



tends to construct small, evenly sized clusters

tends to construct chains of clusters

Images taken from Jonathan Taylor (2010)

Example #4: Clustered Data

(Domain: Data, Elements: Items in Data Set, Cardinality: n-ary)

- Consensus Clustering
  - NP complete
- Heuristics:
  - **Quantitative/metric-based: CSPA**
    Cluster-based Similarity Partitioning Algorithm
  - **Structural/graph-based: HGPA**
    Hyper-Graph Partitioning Algorithm

Example #5: Matchmaker/VisBricks

(Domain: Data, Elements: Clusters in Data Set, Cardinality: binary)

- Relationship: two clusters are related, iff they share data items
- Common visual representation: Ribbons

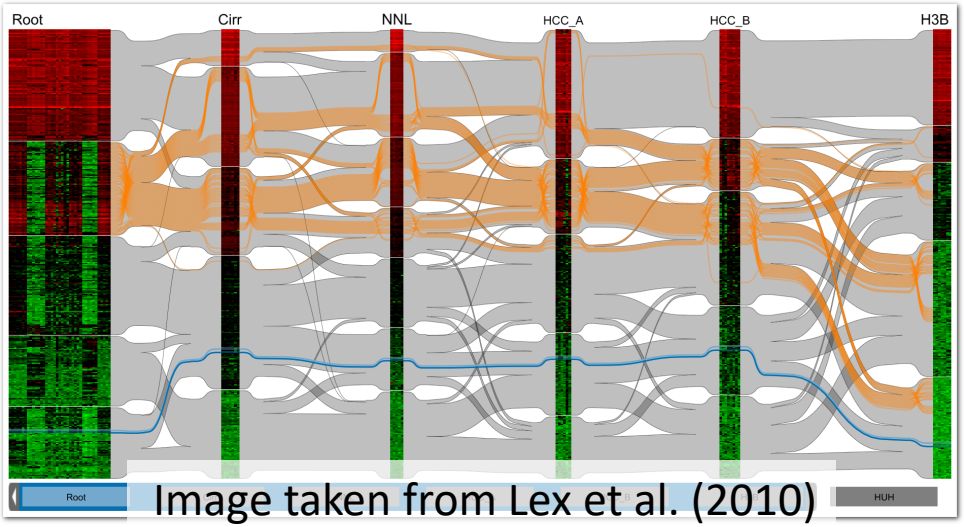


Image taken from Lex et al. (2010)

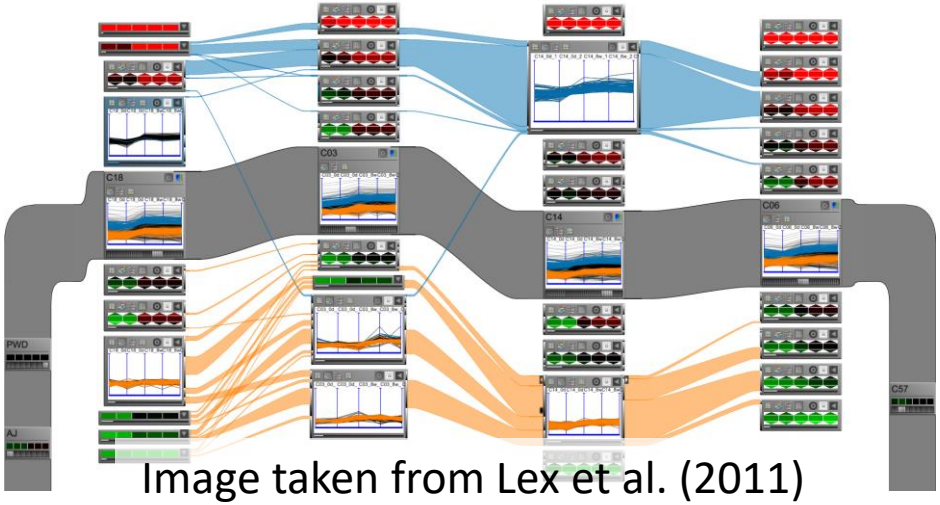


Image taken from Lex et al. (2011)

Example #5: Matchmaker/VisBricks

(Domain: Data, Elements: Clusters in Data Set, Cardinality: binary)

Matchmaker: Clusters of the whole data set

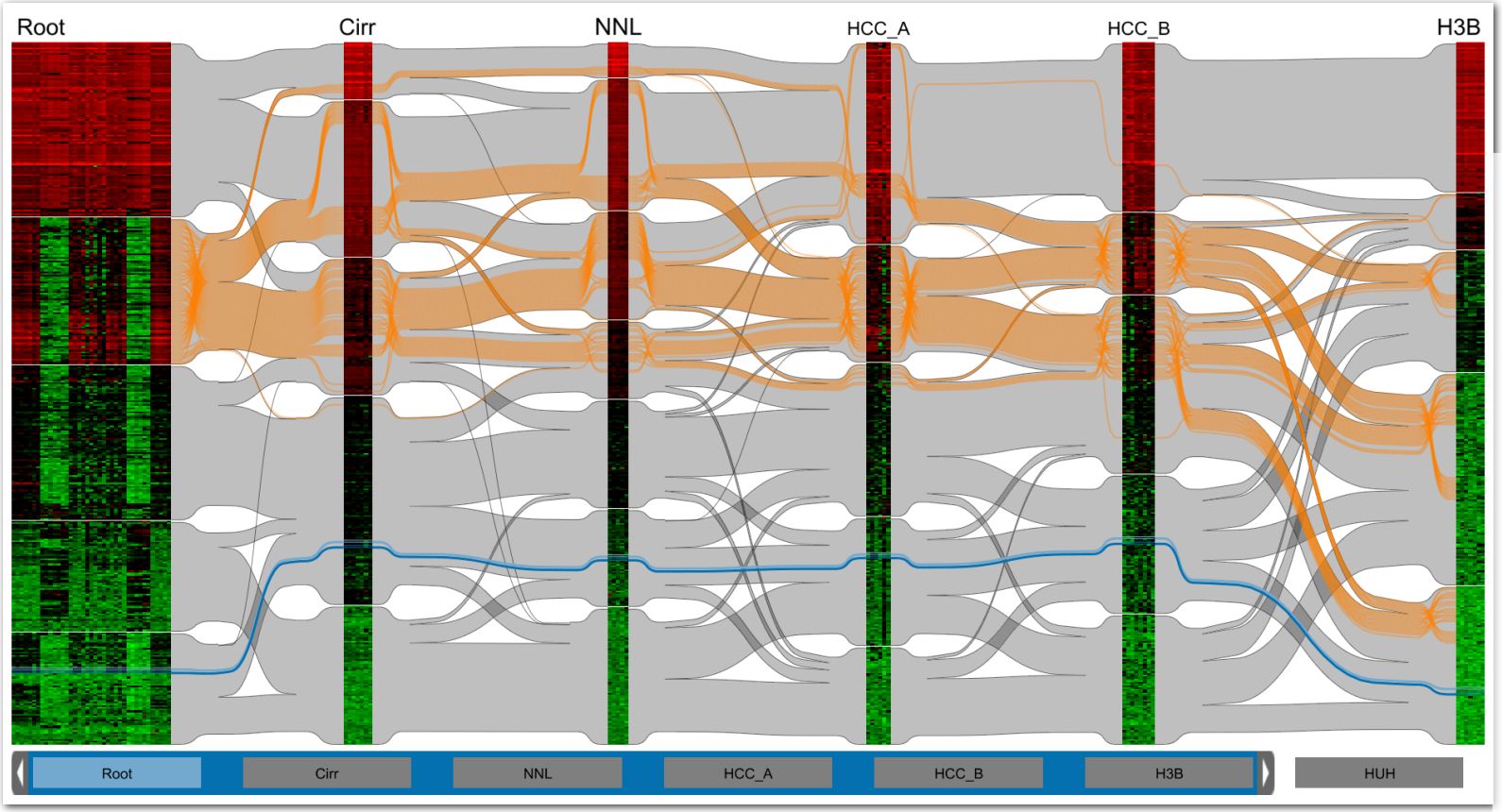


Image taken from Lex et al. (2010)

Example #5: Matchmaker/VisBricks

(Domain: Data, Elements: Clusters in Data Set, Cardinality: binary)

VisBricks: Clusters of dimensional subsets

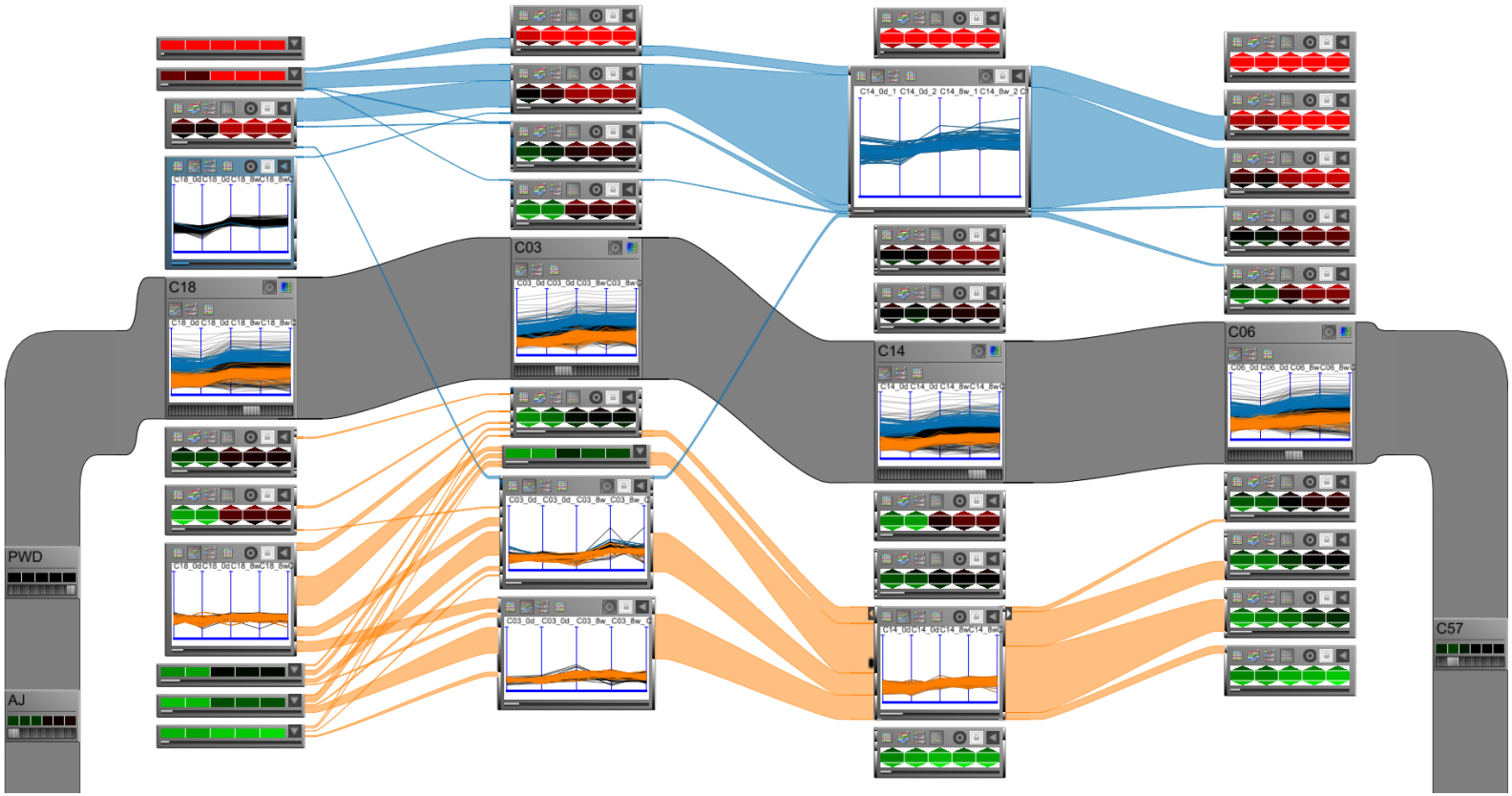


Image taken from Lex et al. (2011)

Example #6: StratomeX

(Domain: Data, Elements: Clusters in Landscape, Cardinality: binary)

- Relationship: two clusters are related, iff they contain data with the same keys (IDs)
- Common visual representation: Ribbons

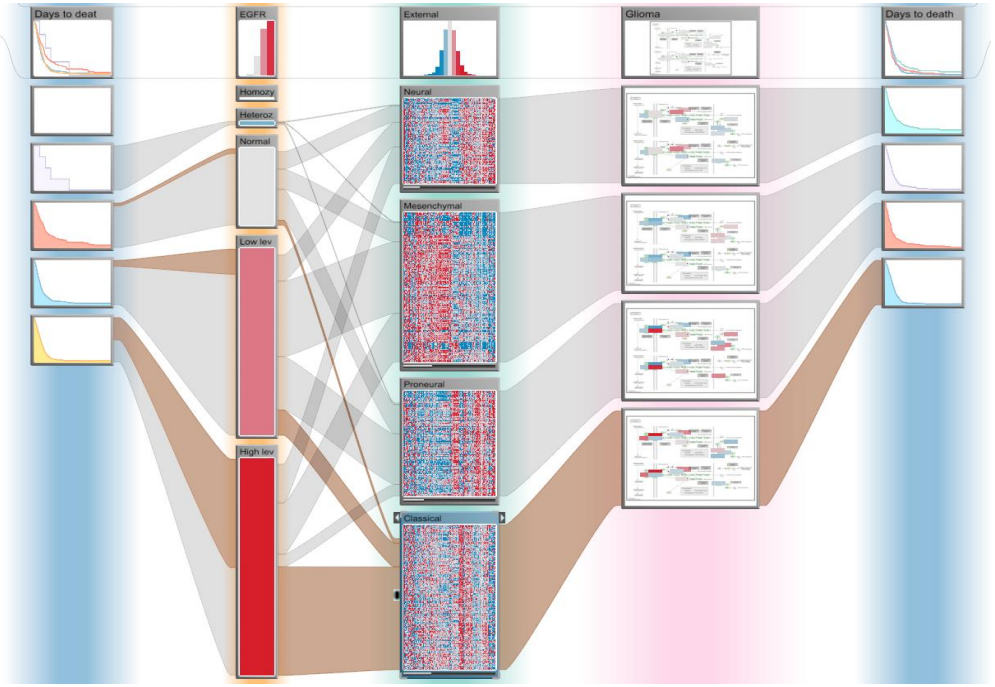


Image taken from Lex et al. (2012)

Example #6: StratomeX

(Domain: Data, Elements: Clusters in Landscape, Cardinality: binary)

- StratomeX: Clusters of different data sets

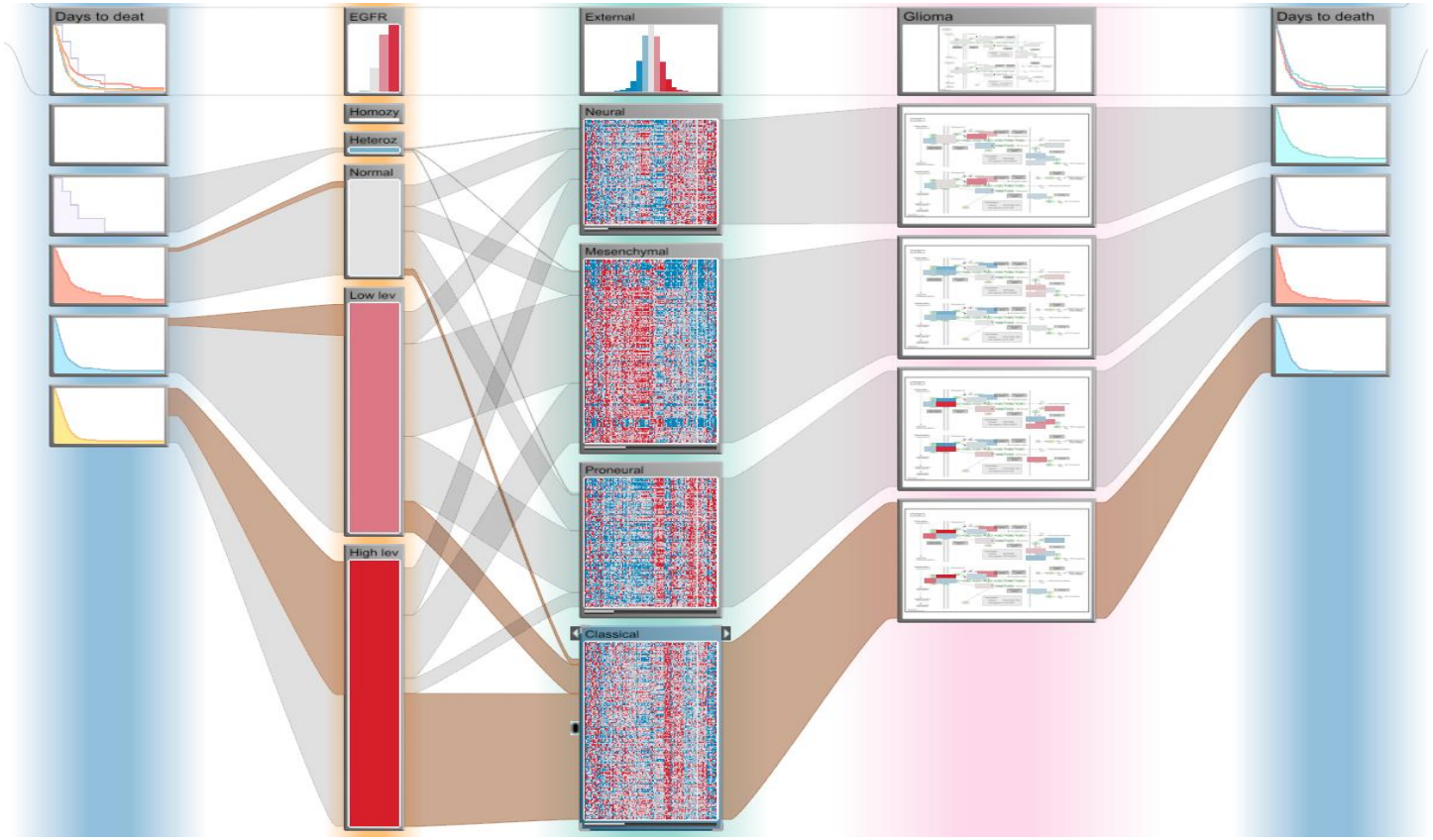


Image taken from Lex et al. (2012)

Example #7: StratomeX DVI

(Domain: Data, Elements: Data Sets in Landscape, Cardinality: binary)

- Relationship: two data sets are related, iff they contain data with the same keys (IDs)
- Common visual representation: Graph

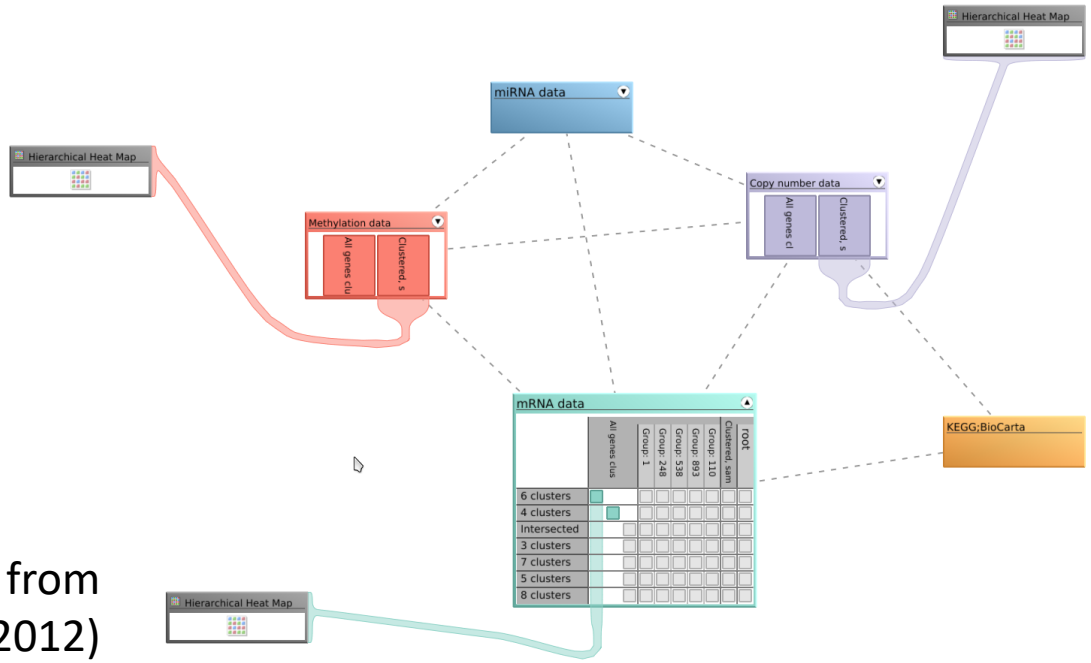
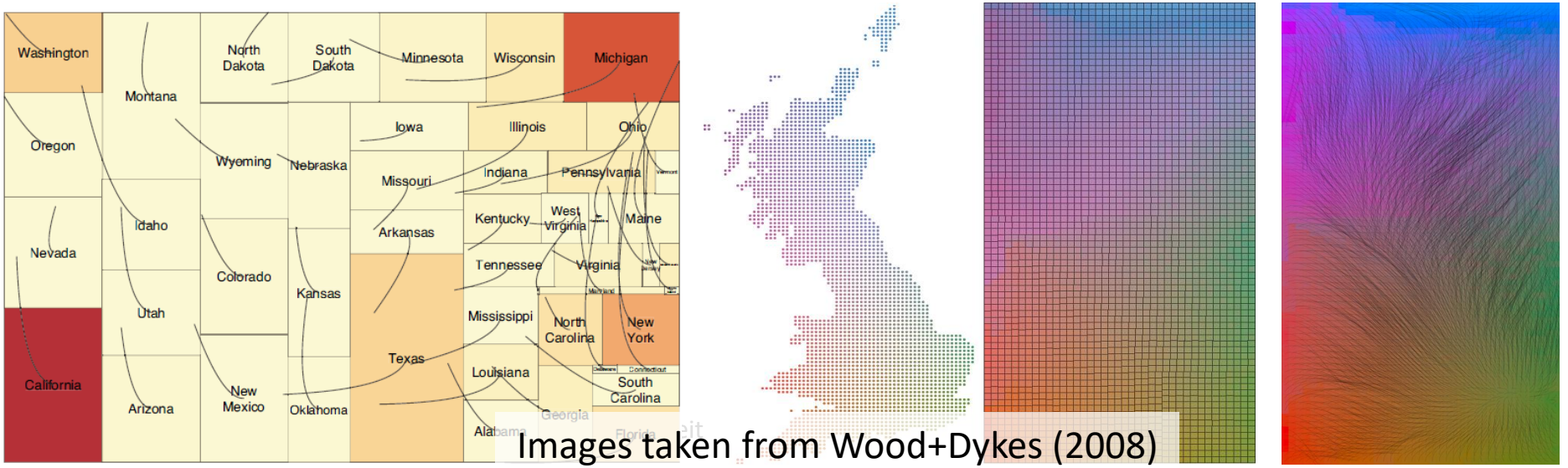


Image taken from Lex et al. (2012)

Example #8: Spatial Treemaps

(Domain: View, Elements: Attributes of Data Items, Cardinality: binary)

- Relationship: two spatial positions are related, iff they both belong to the same data item
- Common visual representation: Strokes and/or Color



Images taken from Wood+Dykes (2008)

Example #8: Spatial Treemaps

(Domain: View, Elements: Attributes of Data Items, Cardinality: binary)

Lines

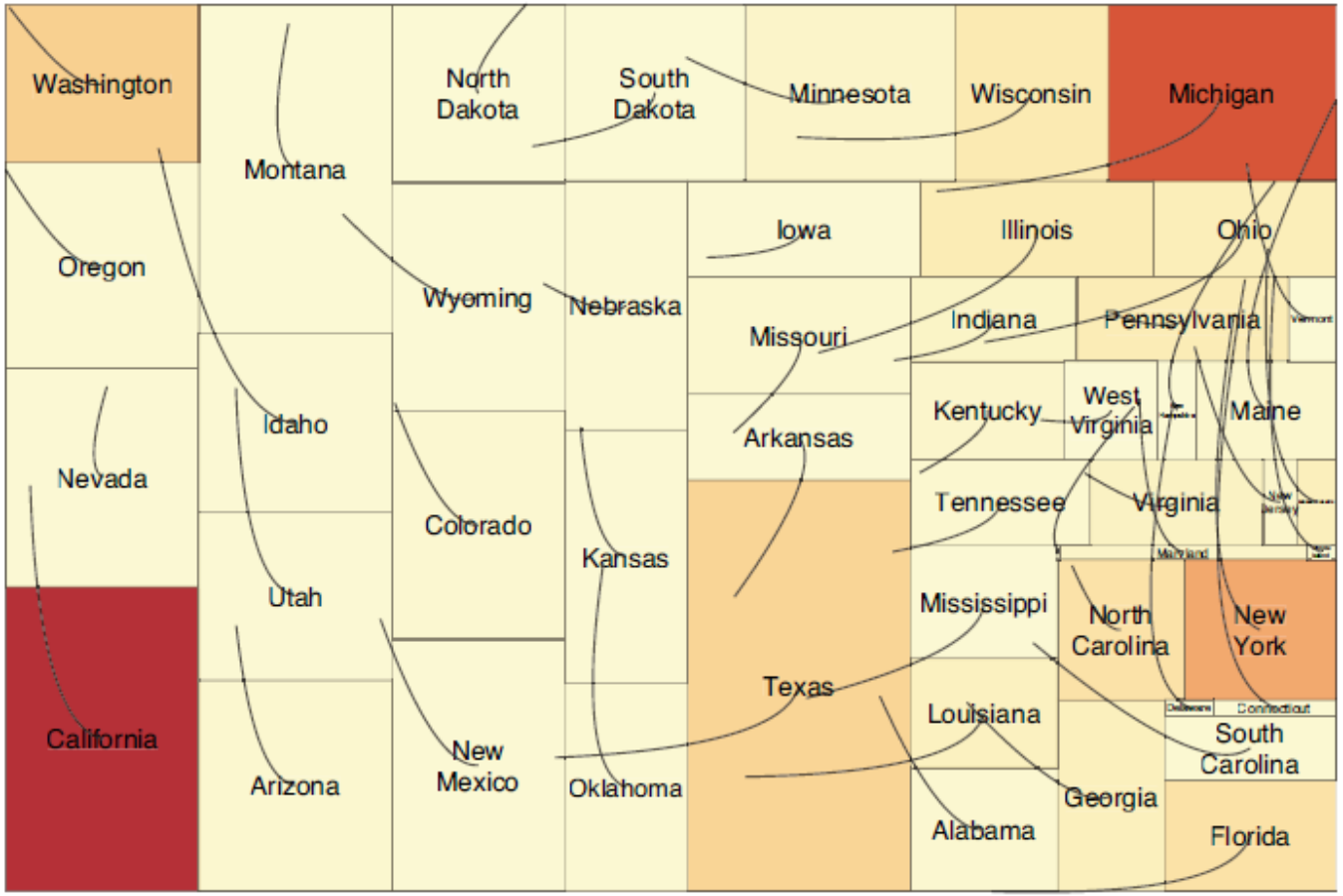
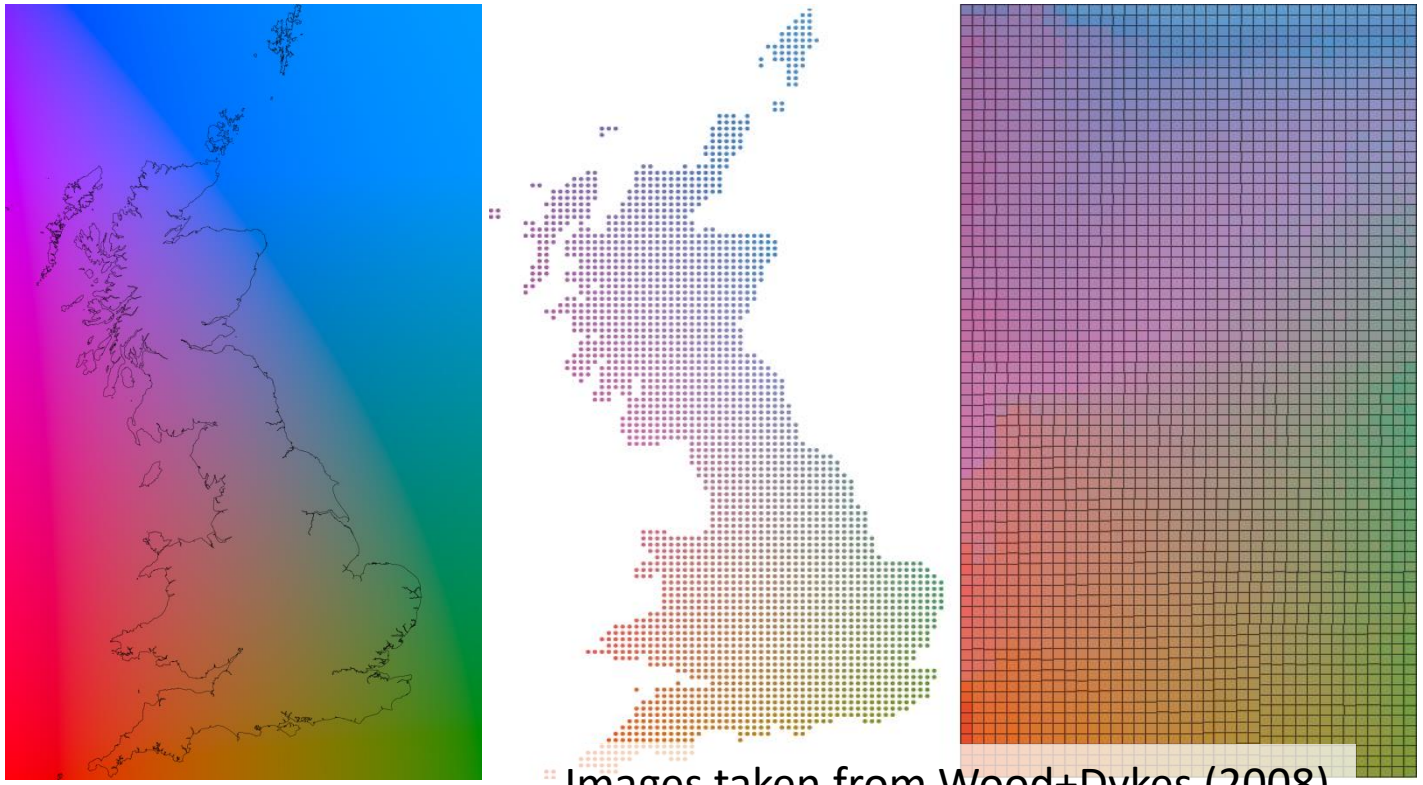


Image taken from Wood+Dykes (2008)

Example #8: Spatial Treemaps

(Domain: View, Elements: Attributes of Data Items, Cardinality: binary)

Color

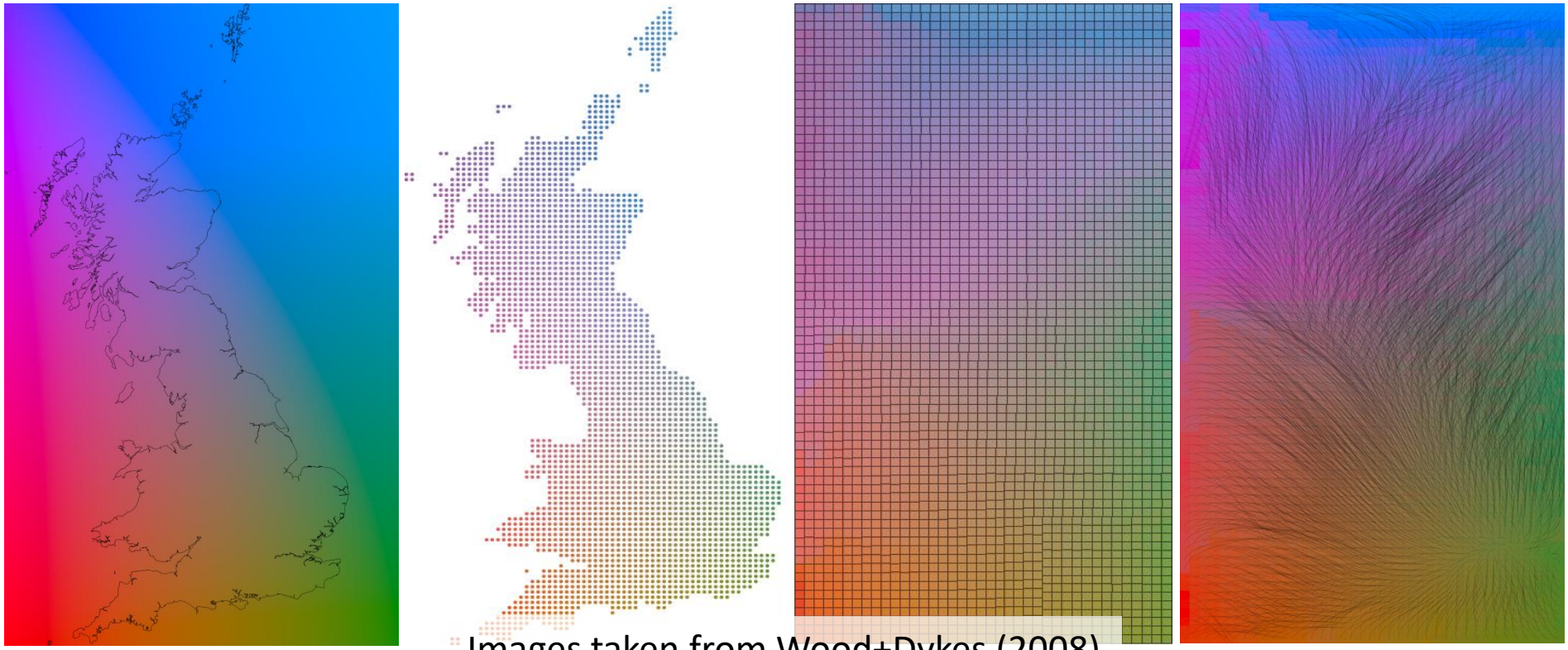


Images taken from Wood+Dykes (2008)

Example #8: Spatial Treemaps

(Domain: View, Elements: Attributes of Data Items, Cardinality: binary)

Color+Lines



Images taken from Wood+Dykes (2008)

Example #9: Stack'n'Flip

(Domain: Interaction, Elements: Data Sets in Landscape, Cardinality: binary)

- Relationship: two data sets are related, iff they are used in sequence
- Common visual representation: Graph

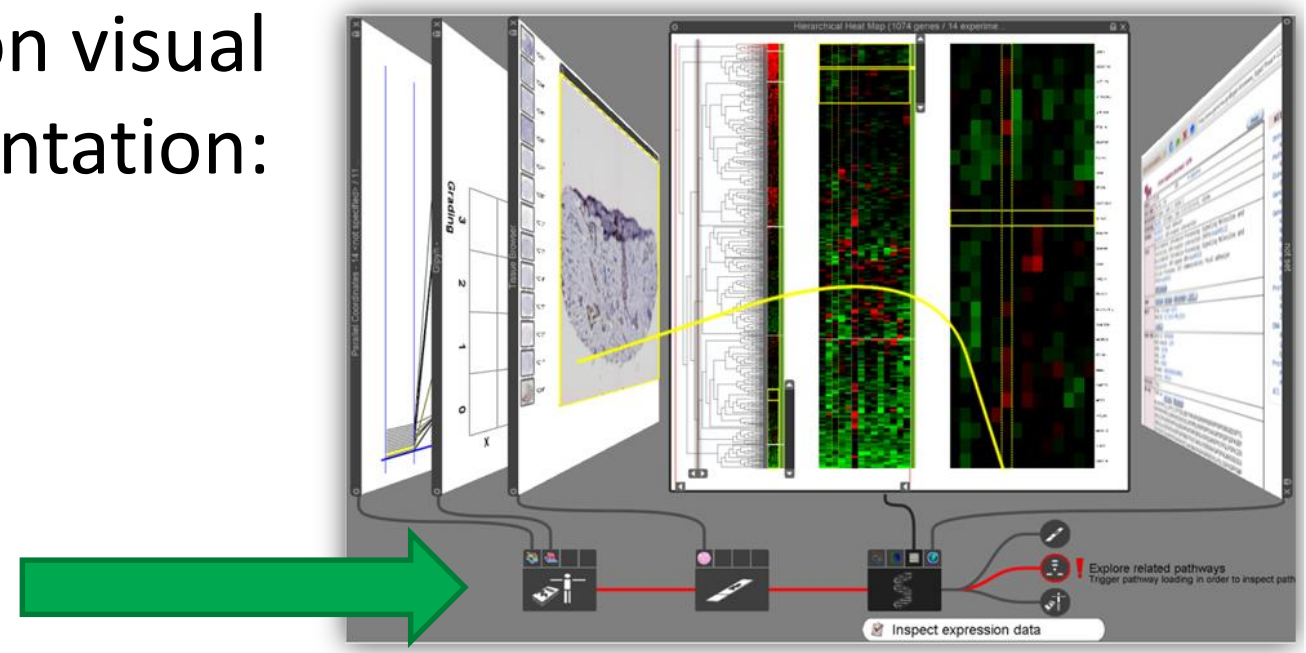
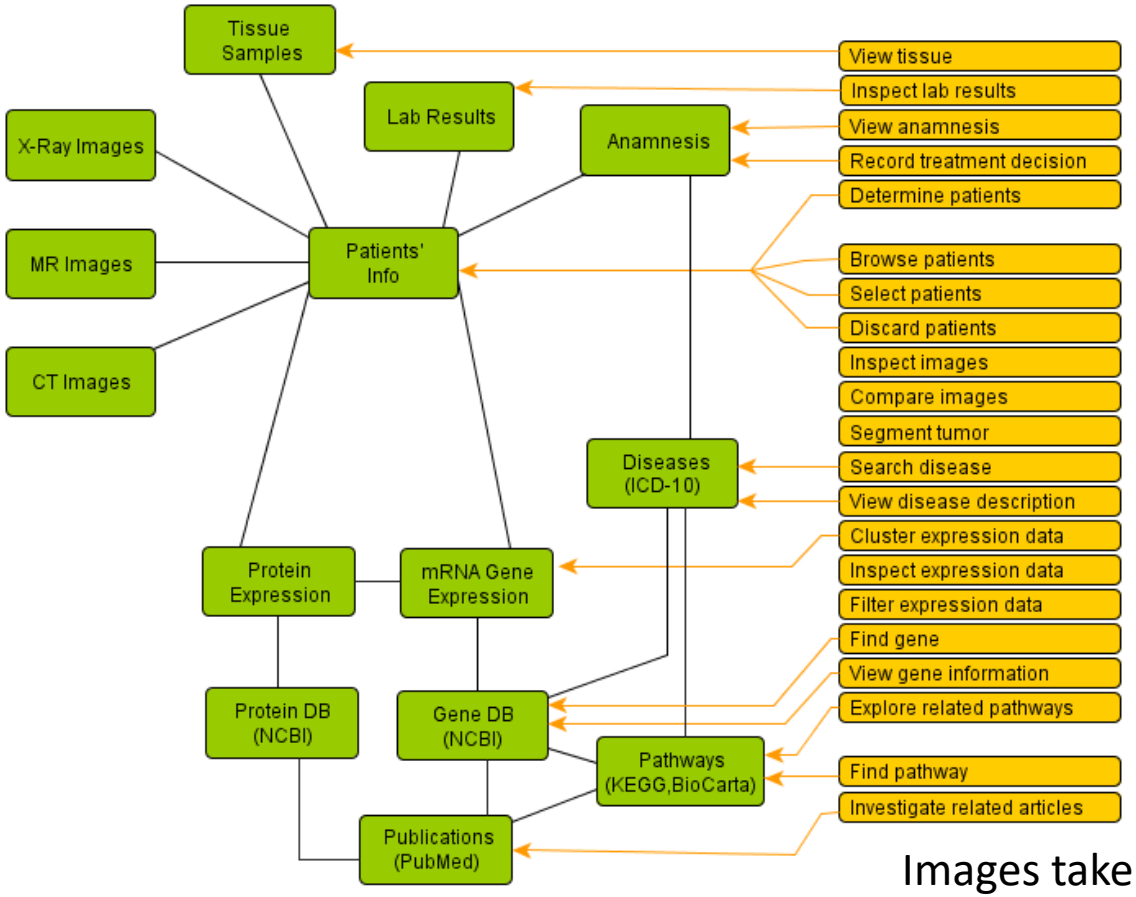
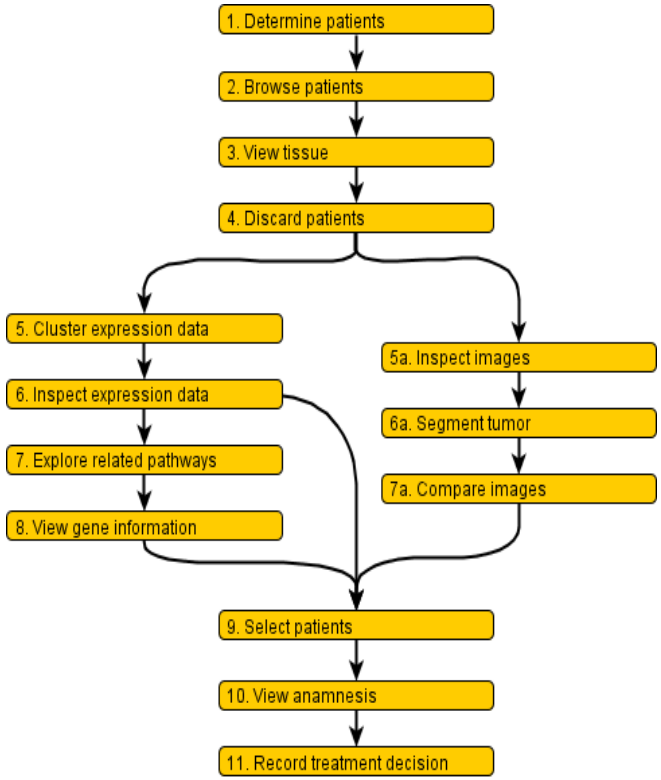


Image taken from Streit et al. (2012)

Example #9: Stack'n'Flip

(Domain: Interaction, Elements: Data Sets in Landscape, Cardinality: binary)





Images taken from Streit et al. (2012)

Example #9: Stack'n'Flip

(Domain: Interaction, Elements: Data Sets in Landscape, Cardinality: binary)

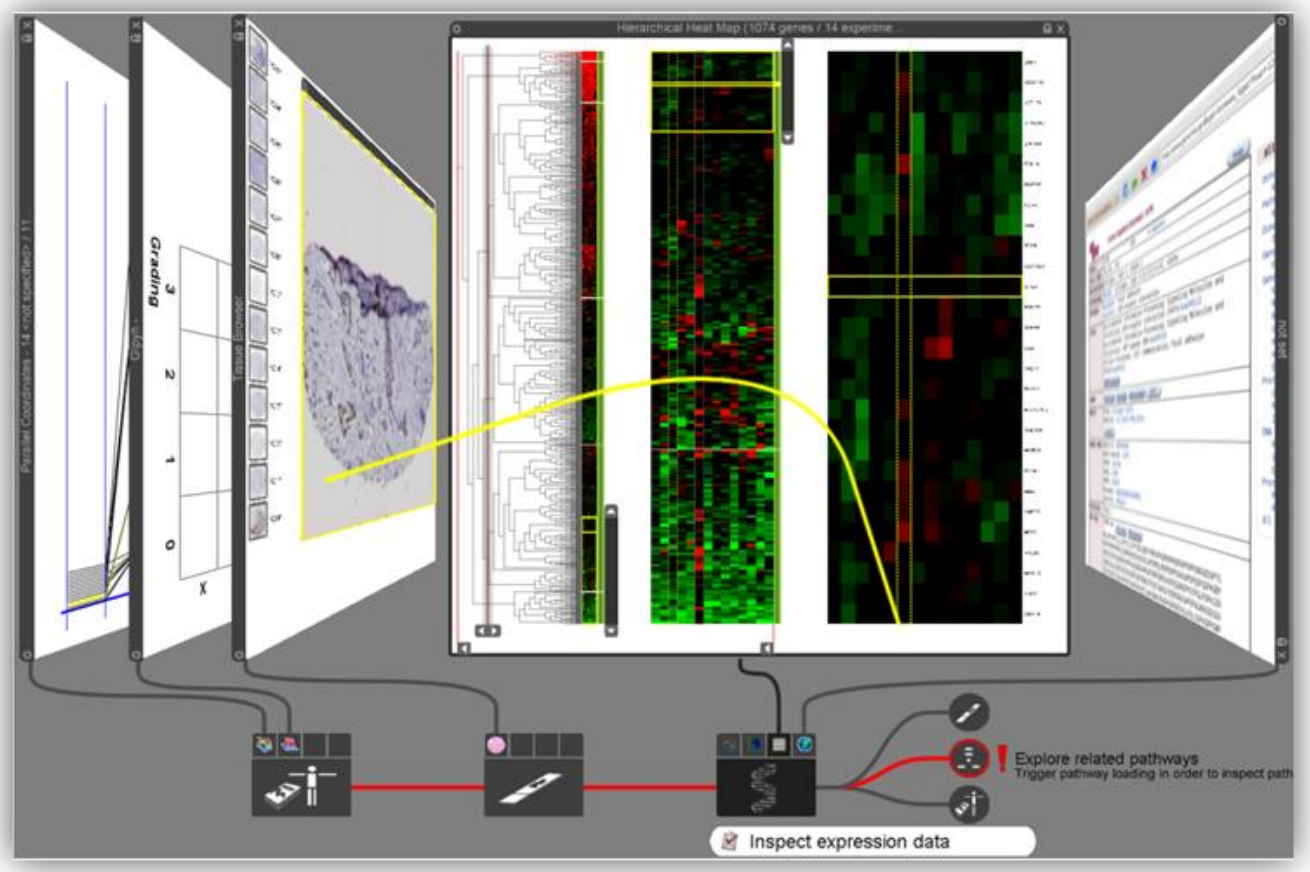


Image taken from Streit et al. (2012)

Edge-based Traveling

Using Relations for Interaction

Clicking an edge with one off-screen end leads automatically towards it

Tooltip indicates details about the off-screen node

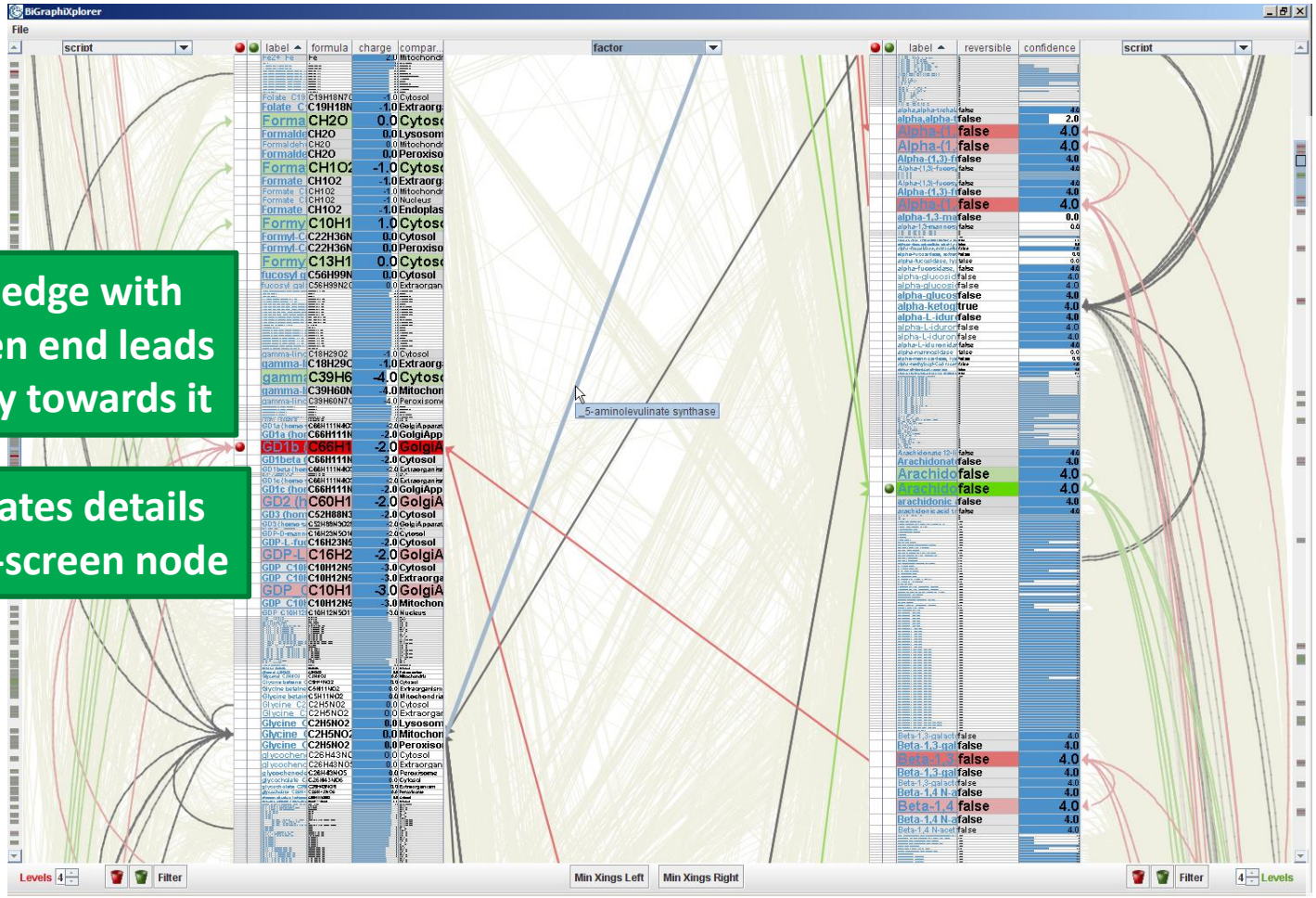


Image taken from Schulz et al. (2008)

What to do with a relation?

- Represent in data space:
  e.g., multidimensional data → graph
  → use standard graph visualization
- Represent in view space:

This is what “Part 2: How to Link” is all about!