Comparative Visualization: Interactive Designs and Algorithms Depending on Data and Tasks

Tatiana von Landesberger[1], Kathrin Ballweg[1], Hans-Jörg Schulz[2], Natalie Kerracher[3], Margit Pohl[4]

VIS Tutorial 2018

1. TU Darmstadt, Darmstadt, Germany
2. Aarhus University, Denmark
3. Edinburgh Napier University, UK
4. TU Wien, Austria

PART I:
THE COMPARISON PROBLEM

Comparison by Data, Task, and Type

Variable width chart — Two variables per item

Table or tables with embedded charts — Many categories

Bar chart horizontal / **Bar chart vertical** — Few categories

Circular area chart — Cyclical data

Line chart — Non-cyclical data

Many periods

Bar chart vertical — Single or few categories

Line chart — Many categories

Few periods

Over time

One variable per item

Among items

Comparison

Relationship

Scatter plot — Two variables

Scatter plot bubble size — Three variables

What would you like to show?

Distribution

Single variable — **Bar histogram** — Few data points

Line histogram — Many data points

Scatter plot — Two variables

Composition

Changing over time

Few periods

Only relative differences matter — **Stacked 100% bar chart**

Relative and absolute differences matter — **Stacked bar chart**

Many periods

Only relative differences matter — **Stacked 100% area chart**

Relative and absolute differences matter — **Stacked area chart**

Static

Simple share of total — **Pie chart**

Accumulation or subtraction to total — **Waterfall chart**

Components of components — **Stacked 100% bar chart w/subcomponents**

Accumulation to total & absolute difference matters — **Tree map**

http://bigdata.black/analytics-predictions/visual-analytics/how-to-choose-the-right-chart/

Chart Chooser Cheat Sheet

Evergreen 2017: Effective Data Visualization: The Right Chart for the Right Data

Why is the problem important?

Only when knowing

- the data to be compared,
- the comparison task to be conducted, and
- the type of the comparison

can we design a visual analysis solution that is

- expressive – i.e., optimally reflects the data,
- effective – i.e., optimally supports the task, and
- appropriate – i.e., scales to the desired type.

http://www.brendangregg.com/blog/2017-02-06/flamegraphs-vs-treemaps-vs-sunburst.html
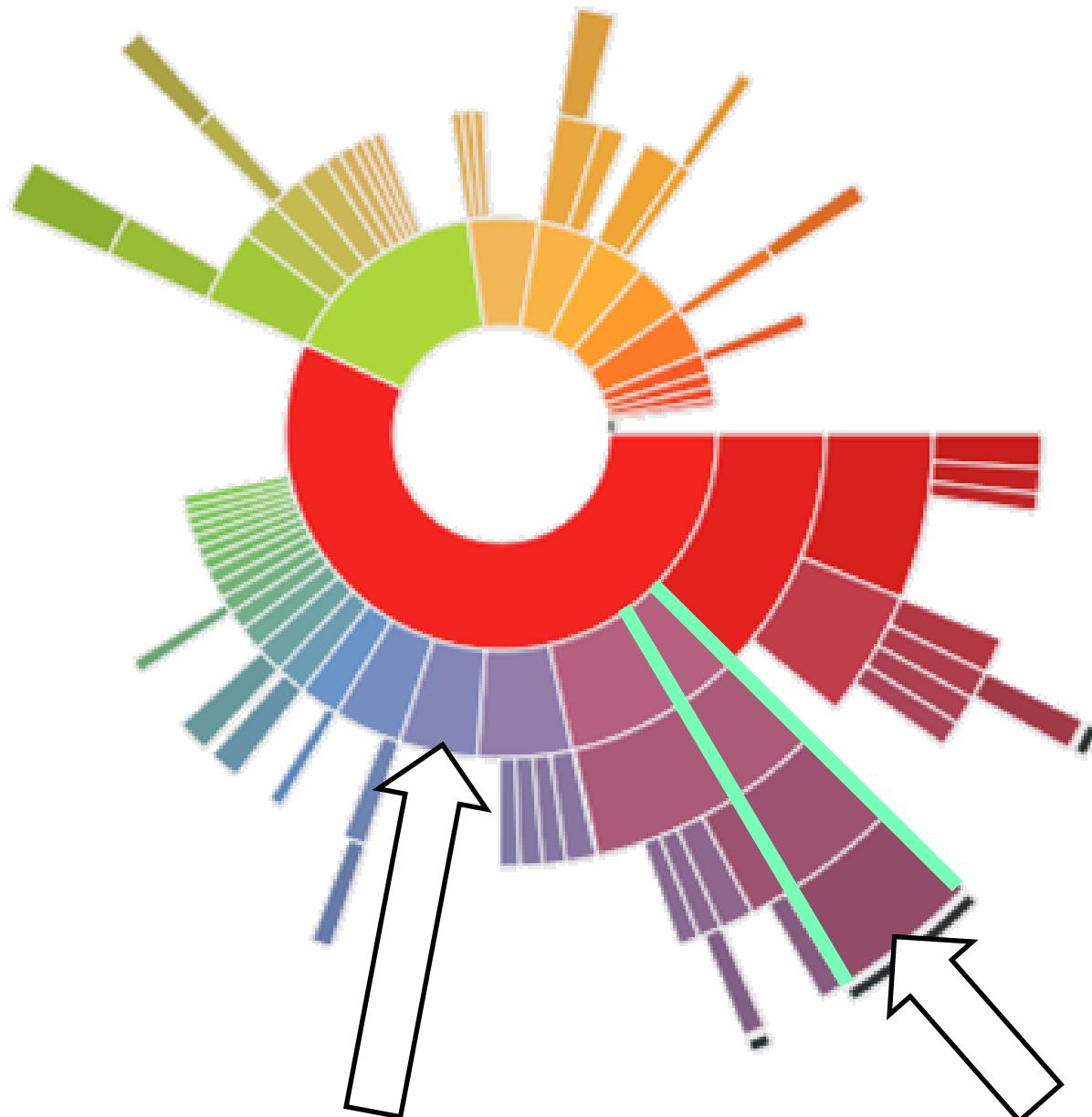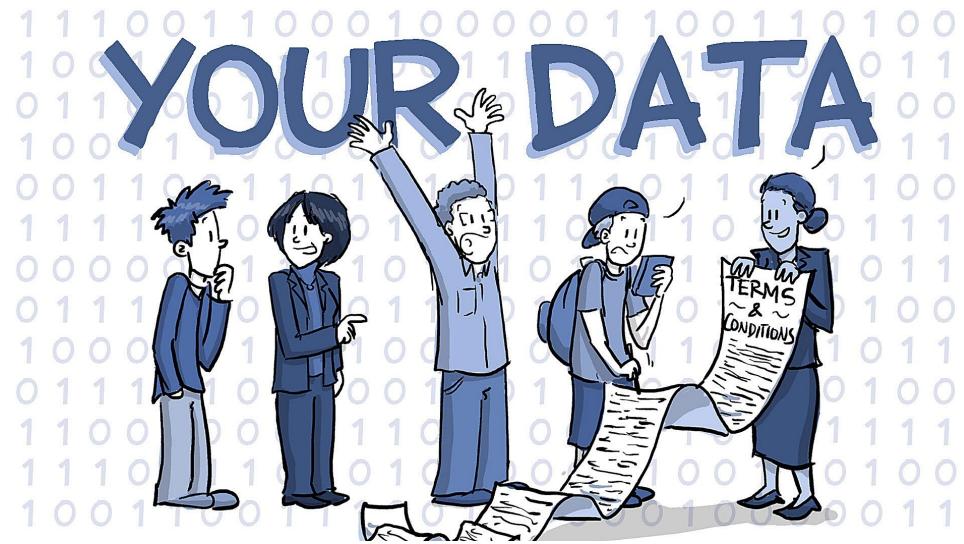
DATA CONSIDERATIONS

Image source: PhD comics, https://youtu.be/y1txYjoSQQc

Data Considerations for Comparison

	Quantitative Data		Qualitative Data	
	Continuous	Discrete	Ordinal	Categorical
Interpolate	✔			
Difference	✔	✔		
Sort	✔	✔	✔	
Match	✔	✔	✔	✔

Adapted from Stevens: On the Theory of Scales of Measurement (1946)

The Stack of Data Comparison

Create Compromise / Composite

$$\phi(x,y)$$

Establish Similarity / Dissimilarity

$$\triangle(x,y)$$

Establish Precedence / Subsequence

$$x \ll y \quad x \gg y$$

Establish Equivalence / Distinctness

$$x \equiv y \quad x \not\equiv y$$

Interpolate

Difference
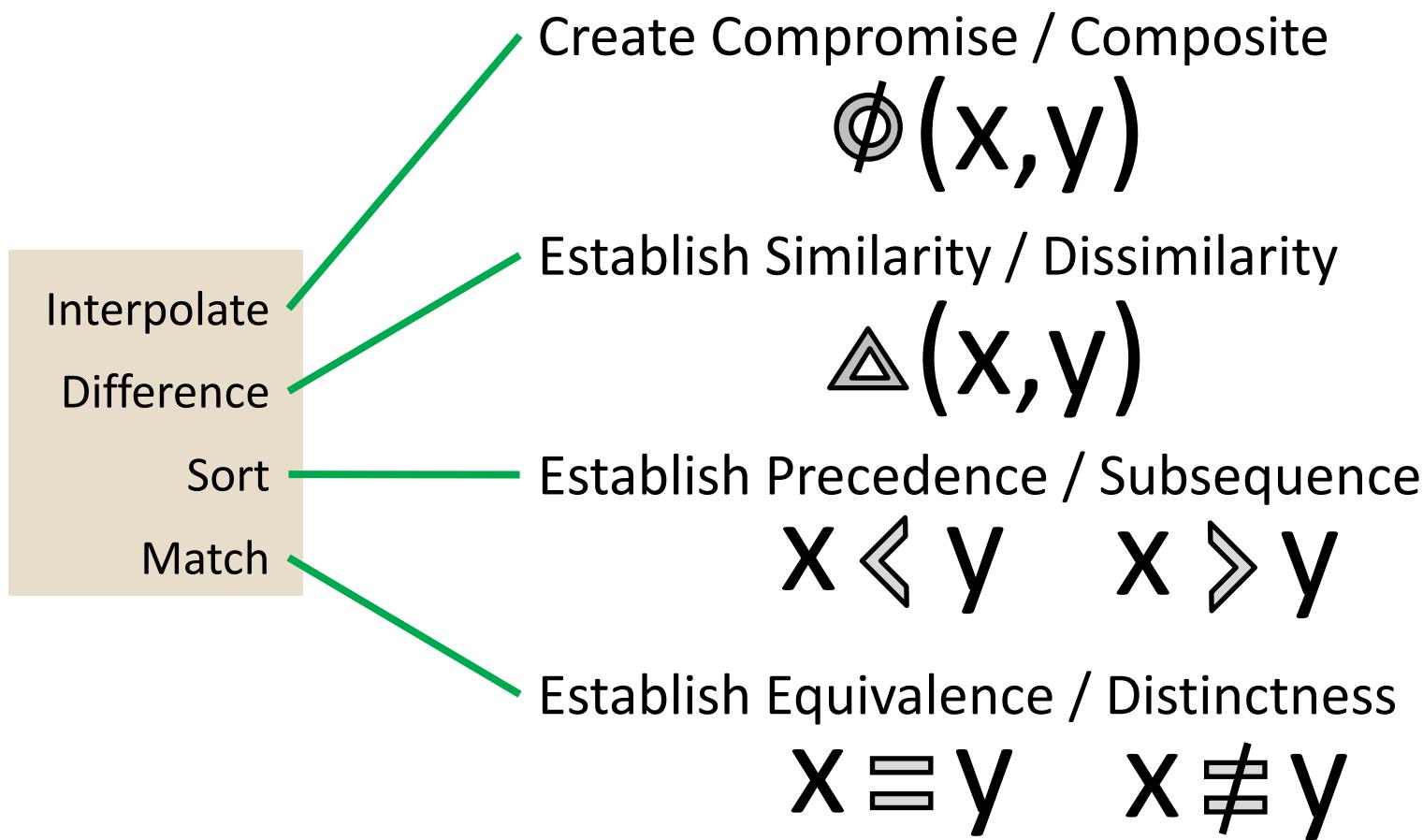
Sort

Match

Distance Metrics for Qualitative Data

Note: Distance ≠ Precedence!

Interpolate

Difference

Sort

Match

Establish Similarity / Dissimilarity

$$\triangle(x,y)$$

- Text / Strings -> Edit Distance
- Graphs / Networks -> Maximal Common Subgraph
- Images -> Pixel by Pixel Color-Comparison in LAB space
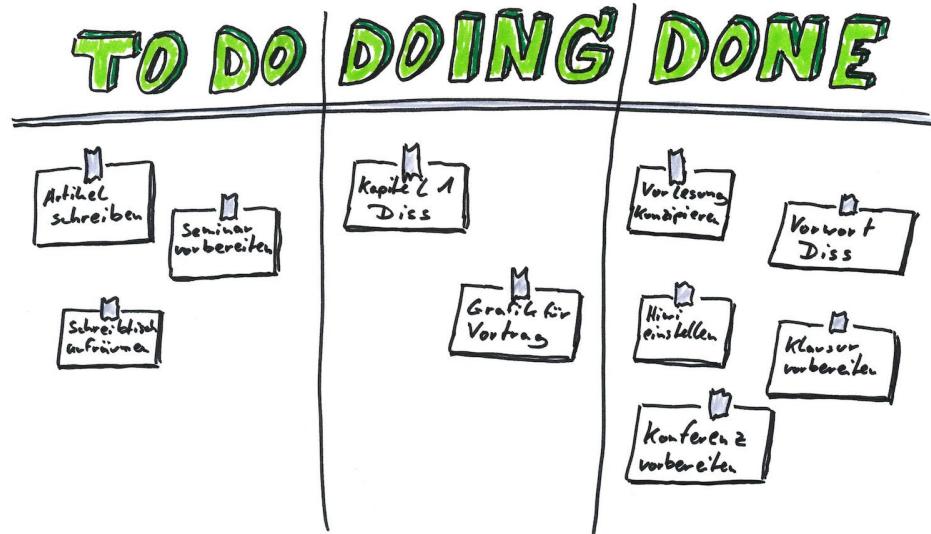
TASK CONSIDERATIONS

Comparison Tasks Deconstructed

- WHY to compare? – Goal (objective/aim/purpose)
   - Exploration
    - Confirmation
    - Presentation

Modeled after [Schulz et al. 2013]

Comparison Goal - Exploration

Comparison as part of an Exploratory Data Analysis:
hypothesis generation through undirected search
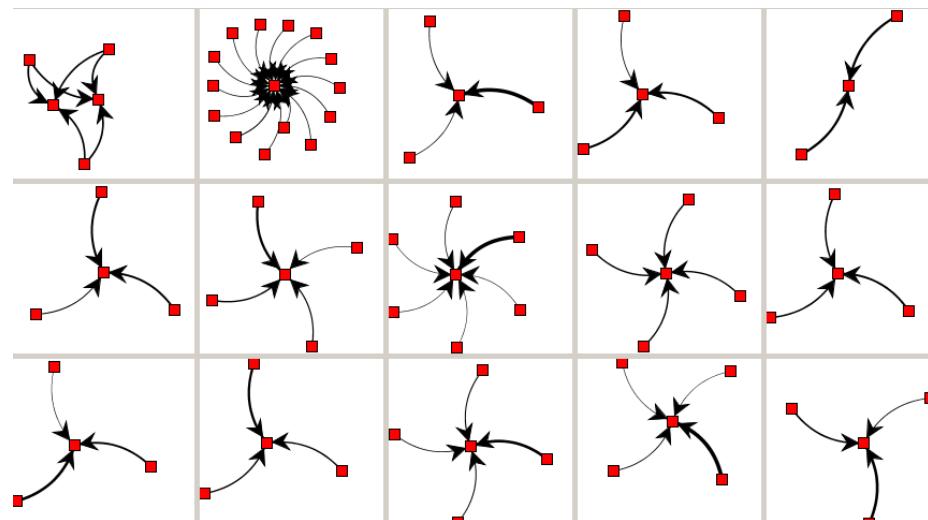
[v.Landesberger et al. 2009]

Comparison Goal - Exploration

Comparison as part of an Exploratory Data Analysis:
hypothesis generation through undirected search
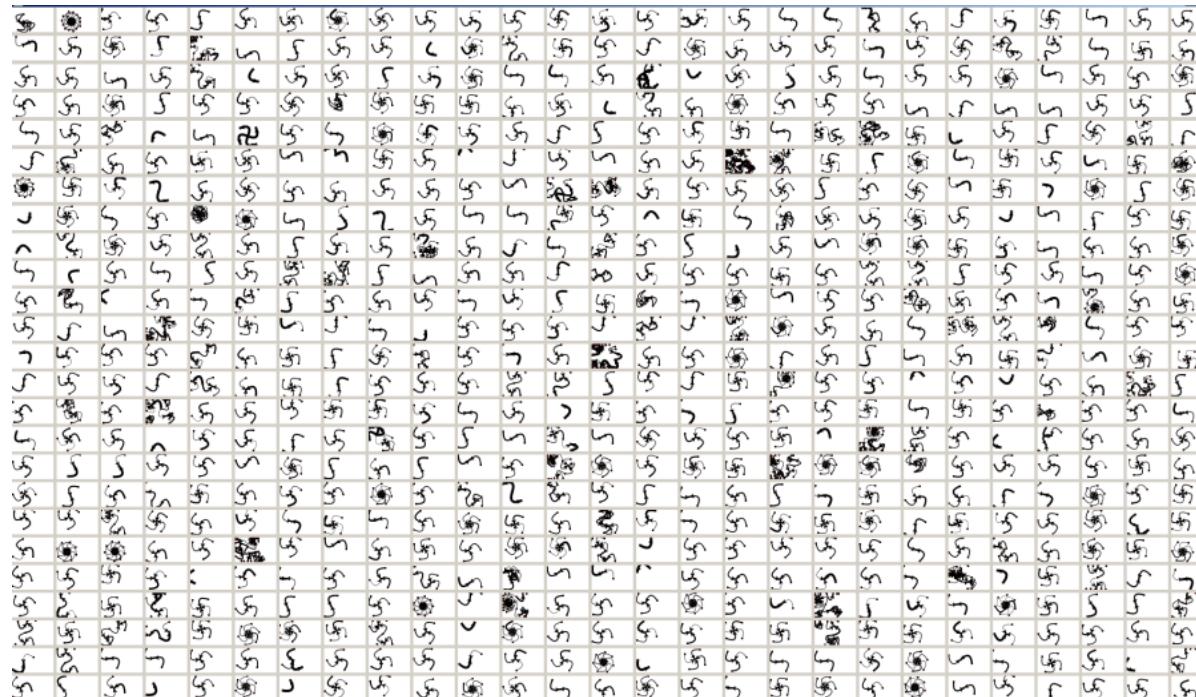
[v.Landesberger et al. 2009]

Comparison Goal - Confirmation

Comparison as part of a Confirmatory Data Analysis:
hypothesis testing through directed search
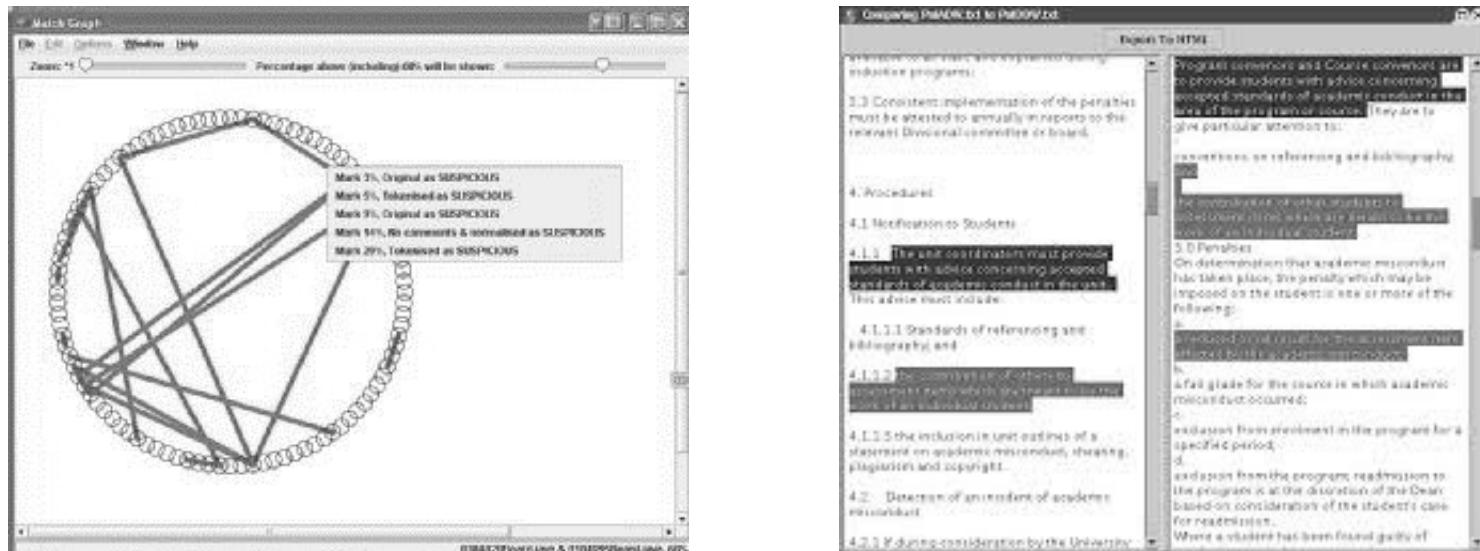
SHERLOCK [White & Joy 2004]

Comparison Goal - Presentation

Comparison as part of Analysis Result Presentation:
communication of similarities

Image source: http://de.guttenplag.wikia.com
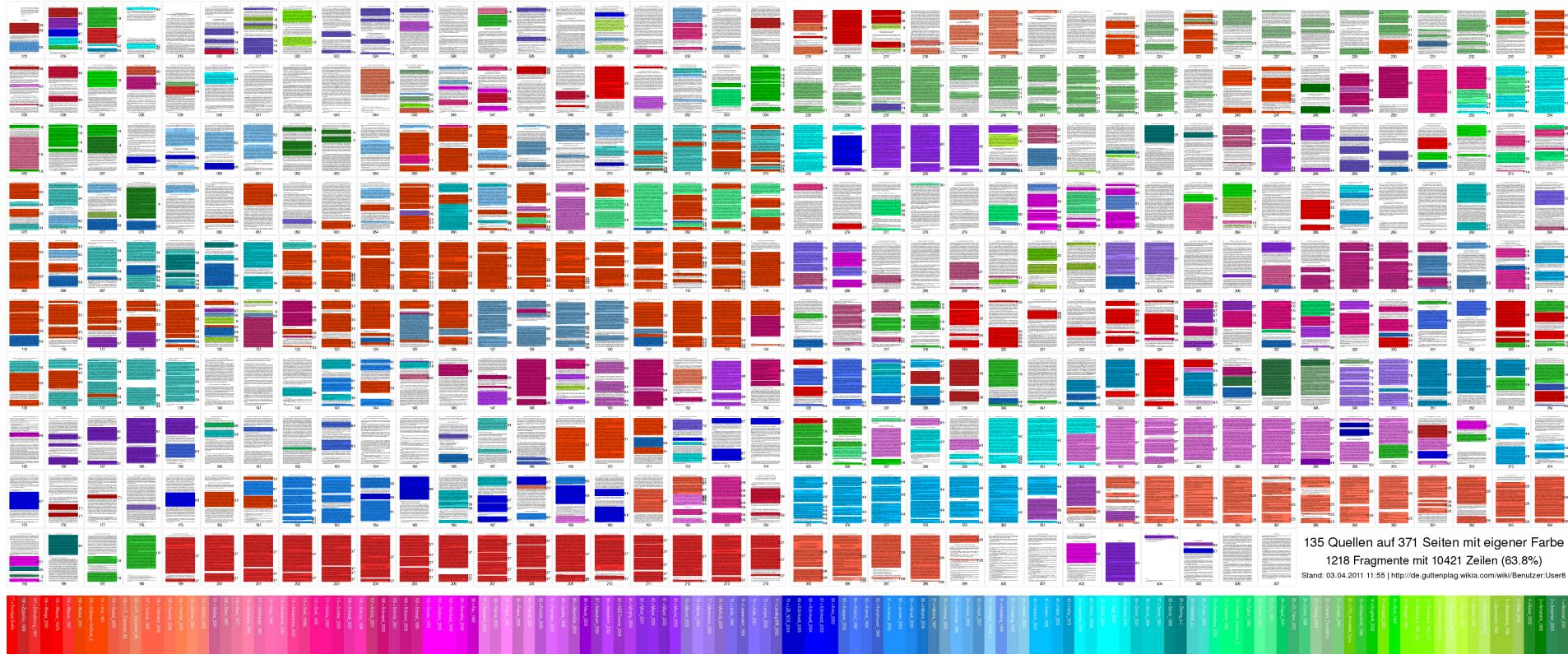
135 Quellen auf 371 Seiten mit eigener Farbe
1218 Fragmente mit 10421 Zeilen (63.8%)
Stand: 03.04.2011 11:55 | http://de.guttenplag.wikia.com/wiki/Benutzer:User8

Comparison Goal - Presentation

Comparison as part of Analysis Result Presentation:
communication of differences

> **70 seats changed hands. The Conservatives lost 13 seats (net) and Labour gained 30. The SNP lost 21.**
>
> | | |
> |---|---|
> | Conservative | -13 (net) |
> | Labour | +30 |
> | SNP | -21 |
> | Lib Dem | +4 |
> | DUP | +2 |
> | Sinn Féin | +3 |
> | Plaid Cymru | +1 |
> | Green | No change |
> | Other | -6 |

Image source: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7979
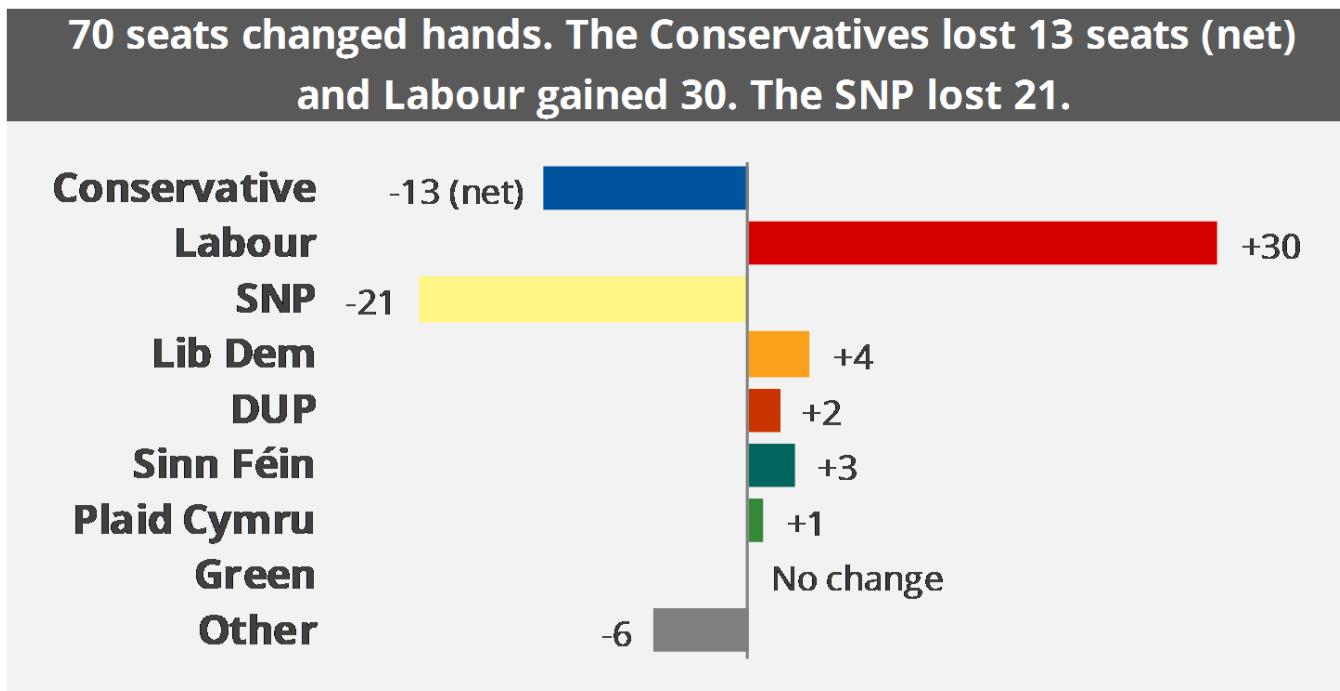
Comparison Goal - Presentation

Comparison through a familiar frame of reference for communicating scale:
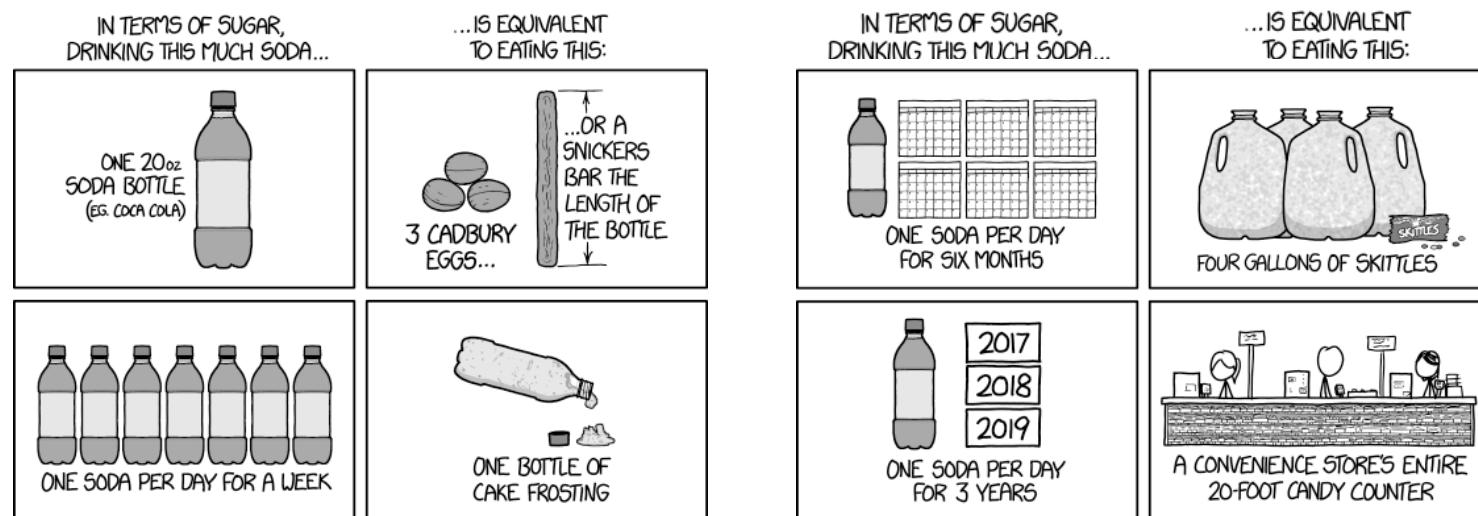
[XKCD 1793]

Comparison Tasks Deconstructed

- WHY to compare? – Goal (objective/aim/purpose)

  - Exploration

  - Confirmation

  - Presentation

- HOW to compare? – Means (action, task)

  - Search / Match

  - Sort / Order

  - Group / Relate

Modeled after [Schulz et al. 2013]

Means of Comparison - Search/Match

Comparison to Template or Exemplar

[Bremm et al. 2011]

Evergreen 2017: Effective Data Visualization: The Right Chart for the Right Data

Means of Comparison - Sort/Order
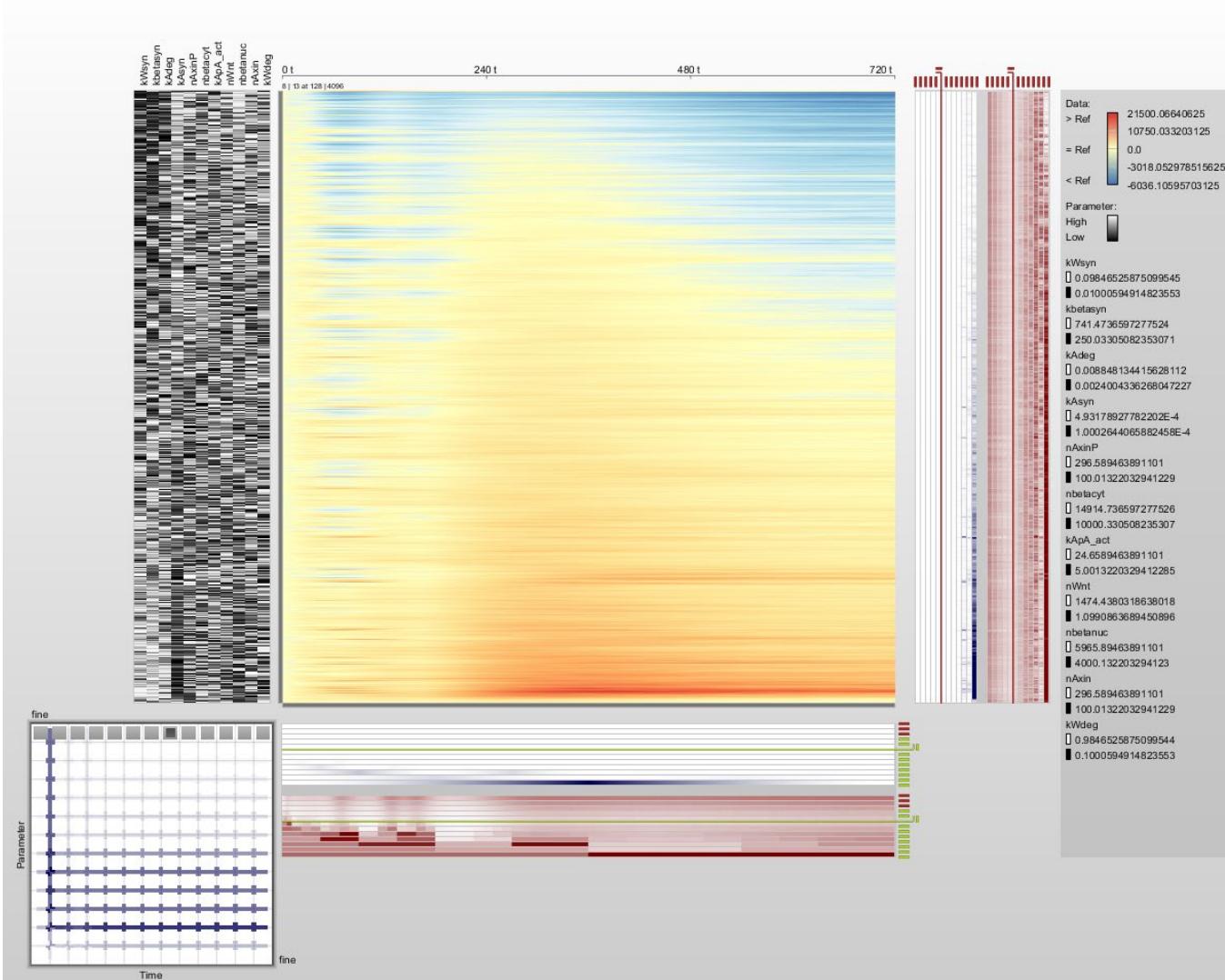
[Luboschik et al. 2014]

Means of Comparison - Group/Relate
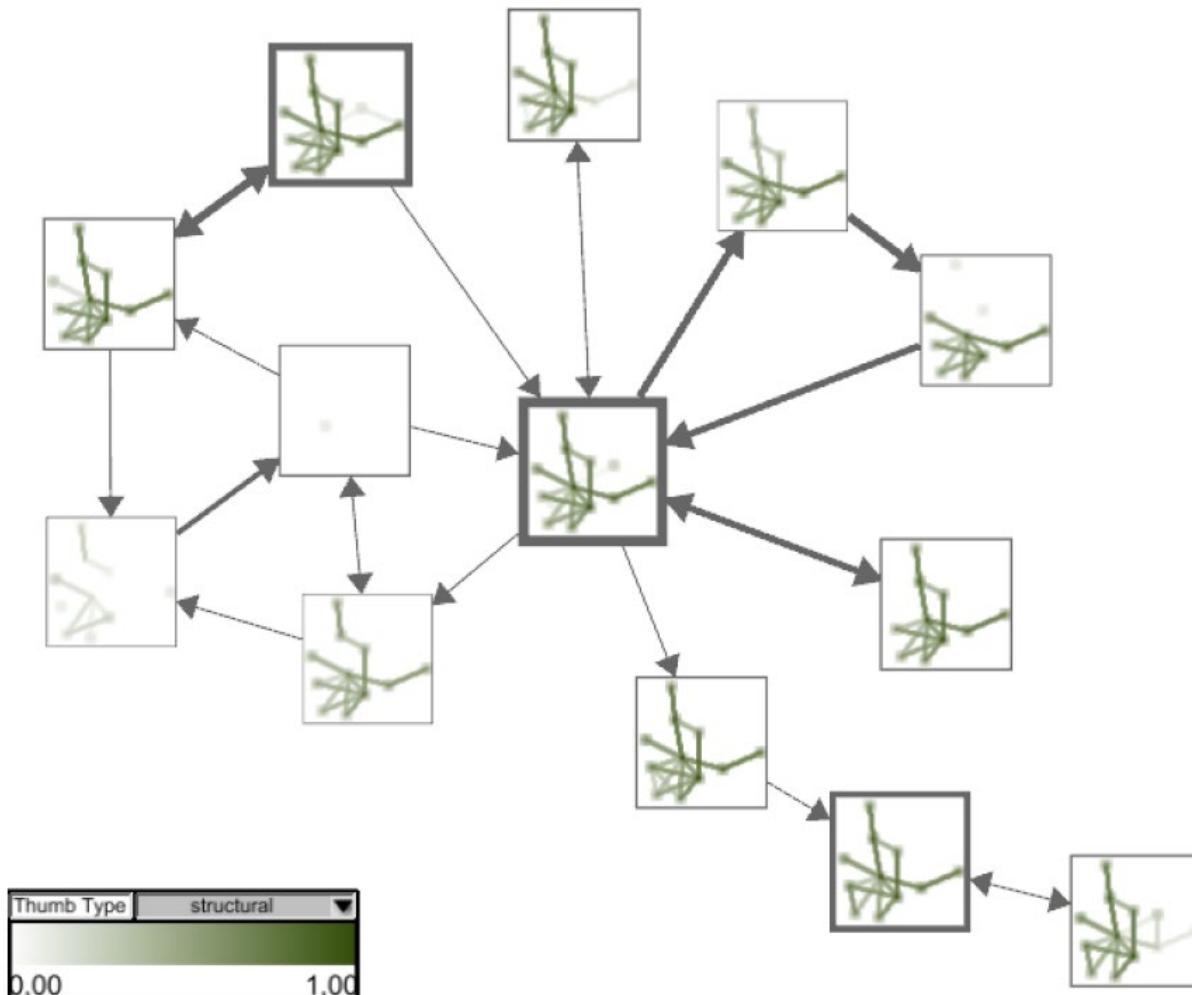
[Hadlak 2015]

Means of Comparison - Group/Relate

https://sctr7.com/2014/08/14/network-analytics-more-than-pretty-pictures/

https://petapixel.com/2016/09/14/20-composition-techniques-will-improve-photos/
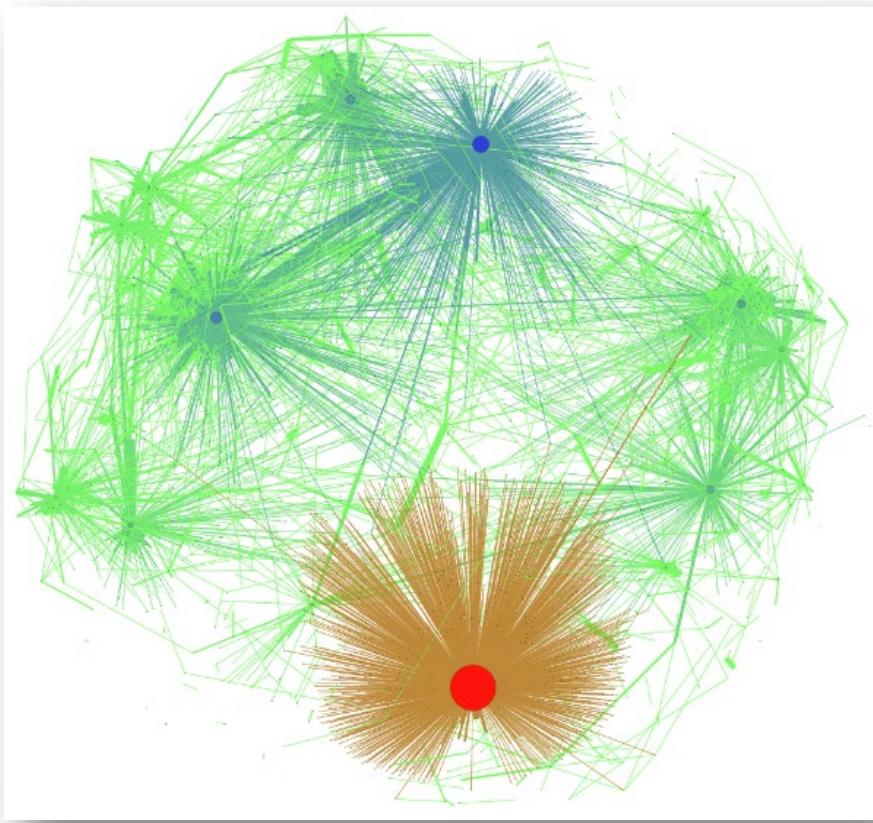
Means of Comparison - Combinations
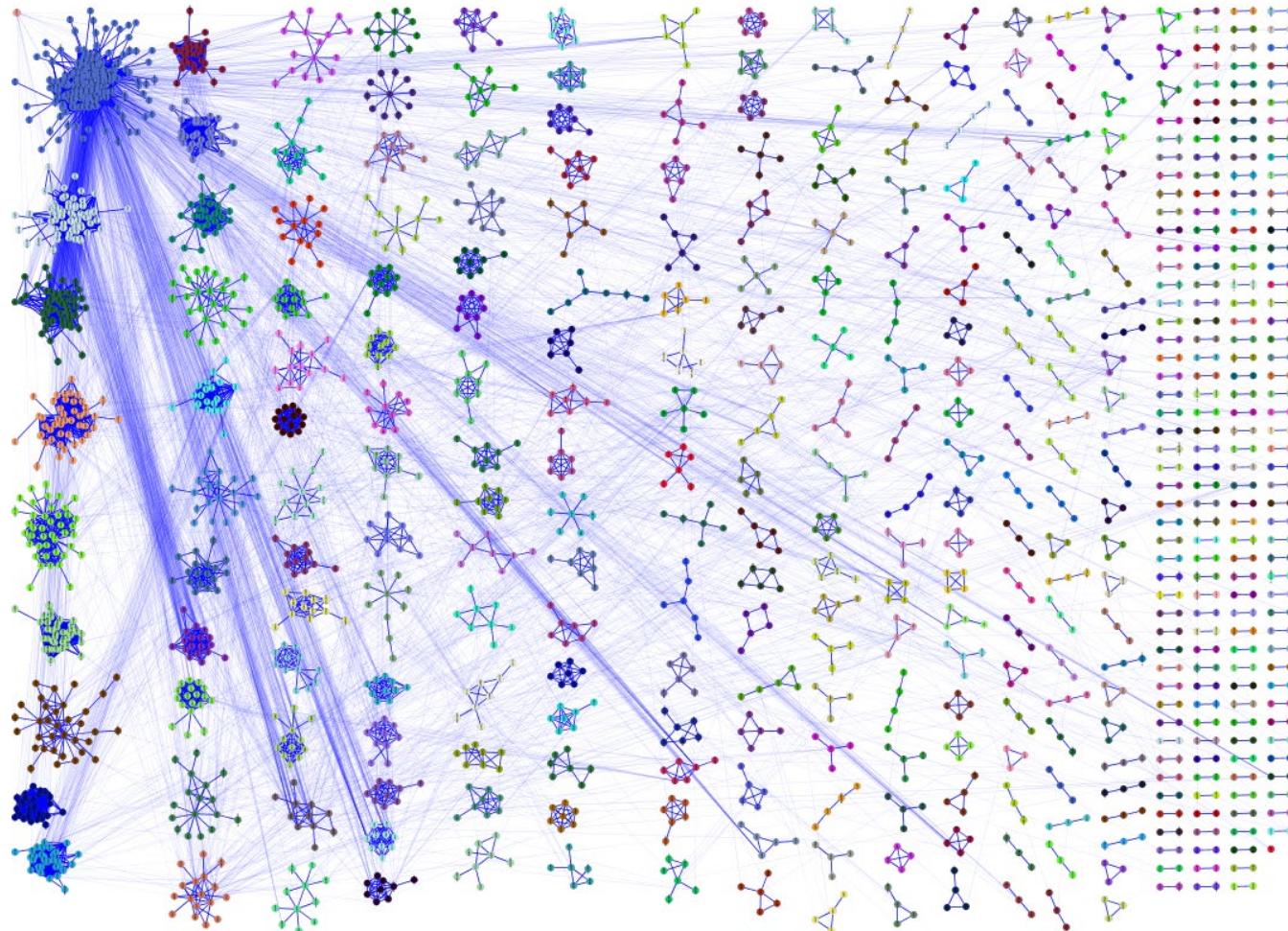
Image Source: http://www.rbvi.ucsf.edu/cytoscape/clusterMaker2/

Comparison Tasks Deconstructed

- WHY to compare? – Goal (objective/aim/purpose)
  - Exploration
  - Confirmation
  - Presentation

- HOW to compare? – Means (action, task)
  - Search / Match
  - Sort / Order
  - Group / Relate

- WHAT to compare? – Data characteristics (data aspect)
  - Data type
 - Data level / granularity
  - Data cardinality

Modeled after [Schulz et al. 2013]

Data Type - Document Collections
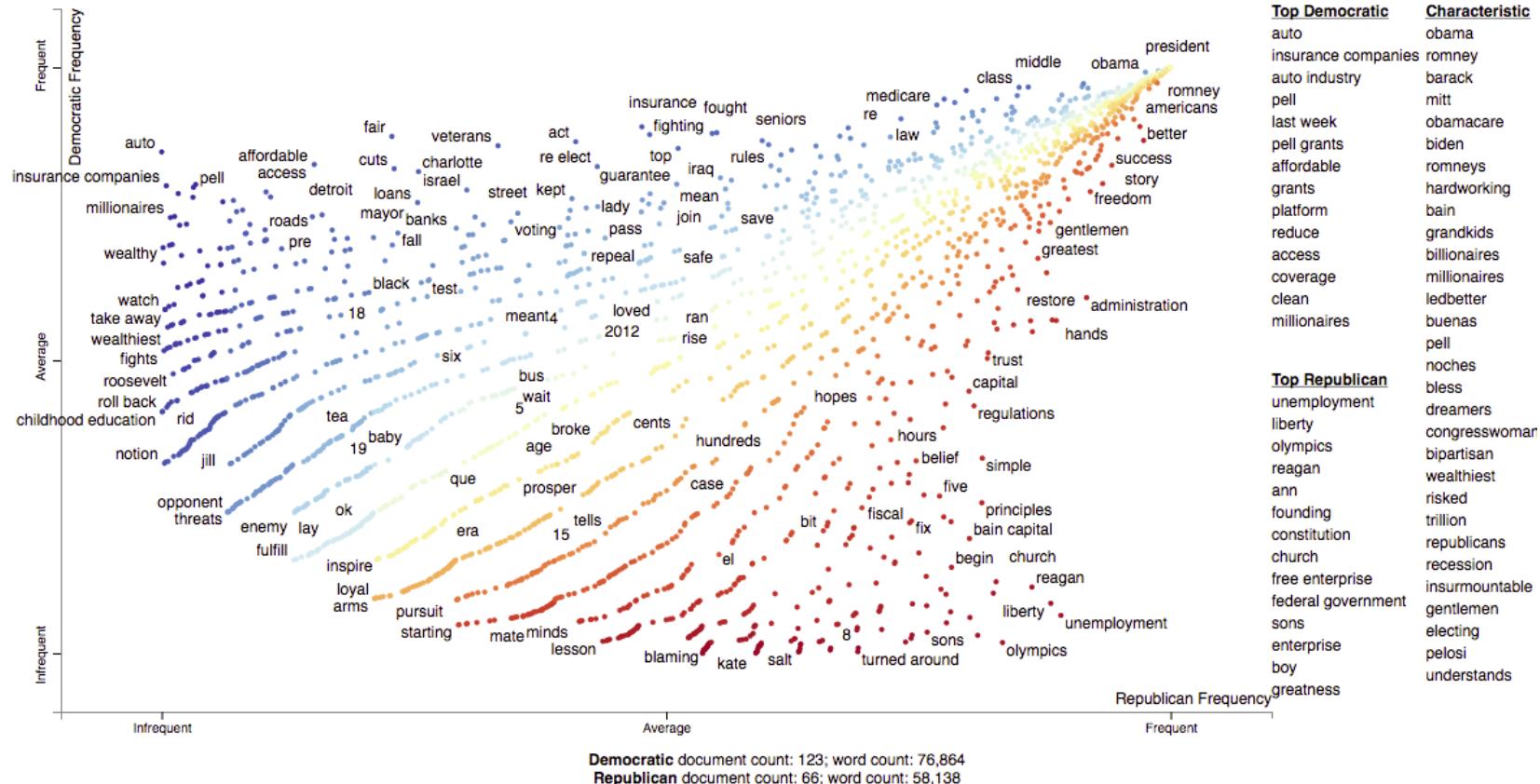
ScatterText - [Kessler 2017]

Data Type - Document Collections

Parallel Tag Clouds - [Collins et al. 2009]

Data Type - Sets
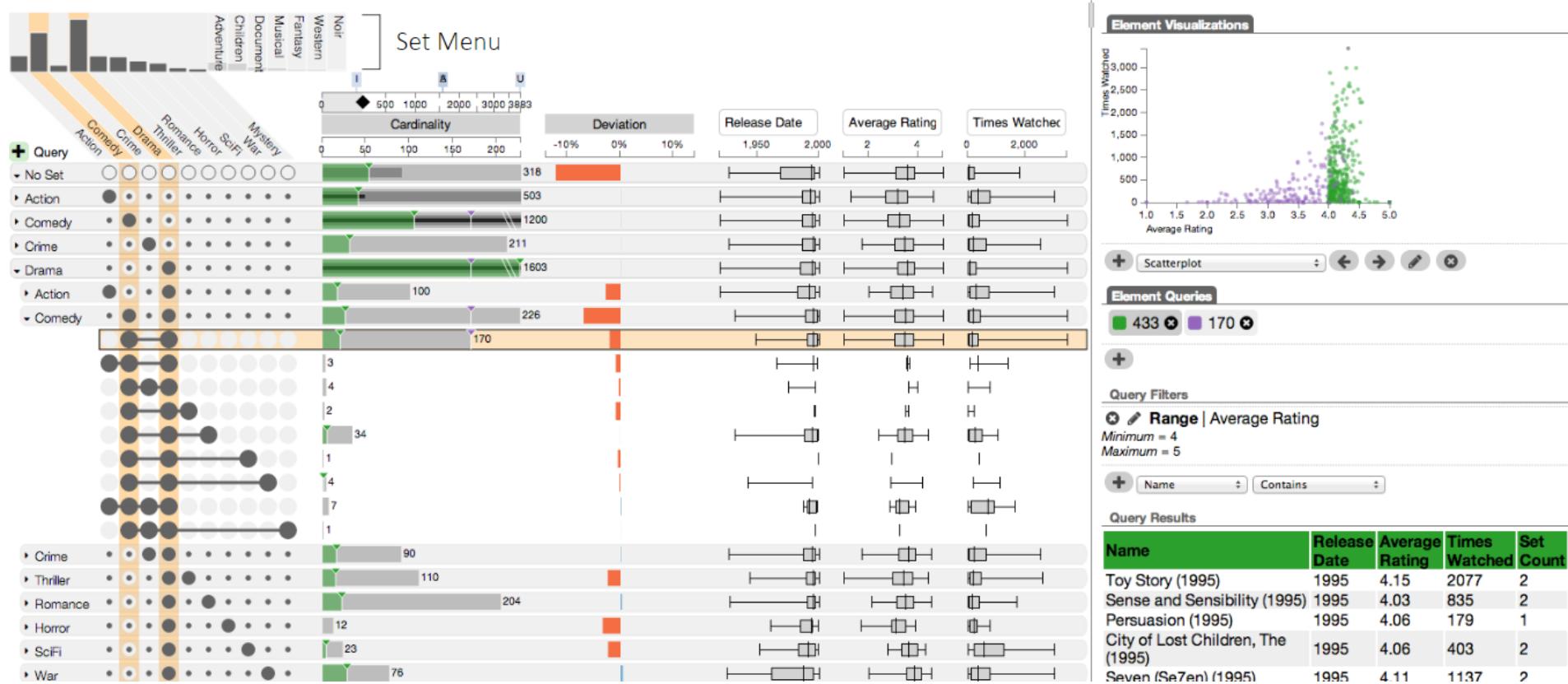
UpSet - [Lex et al. 2014]

Data Type - Sets

UpSet - [Lex et al. 2014]
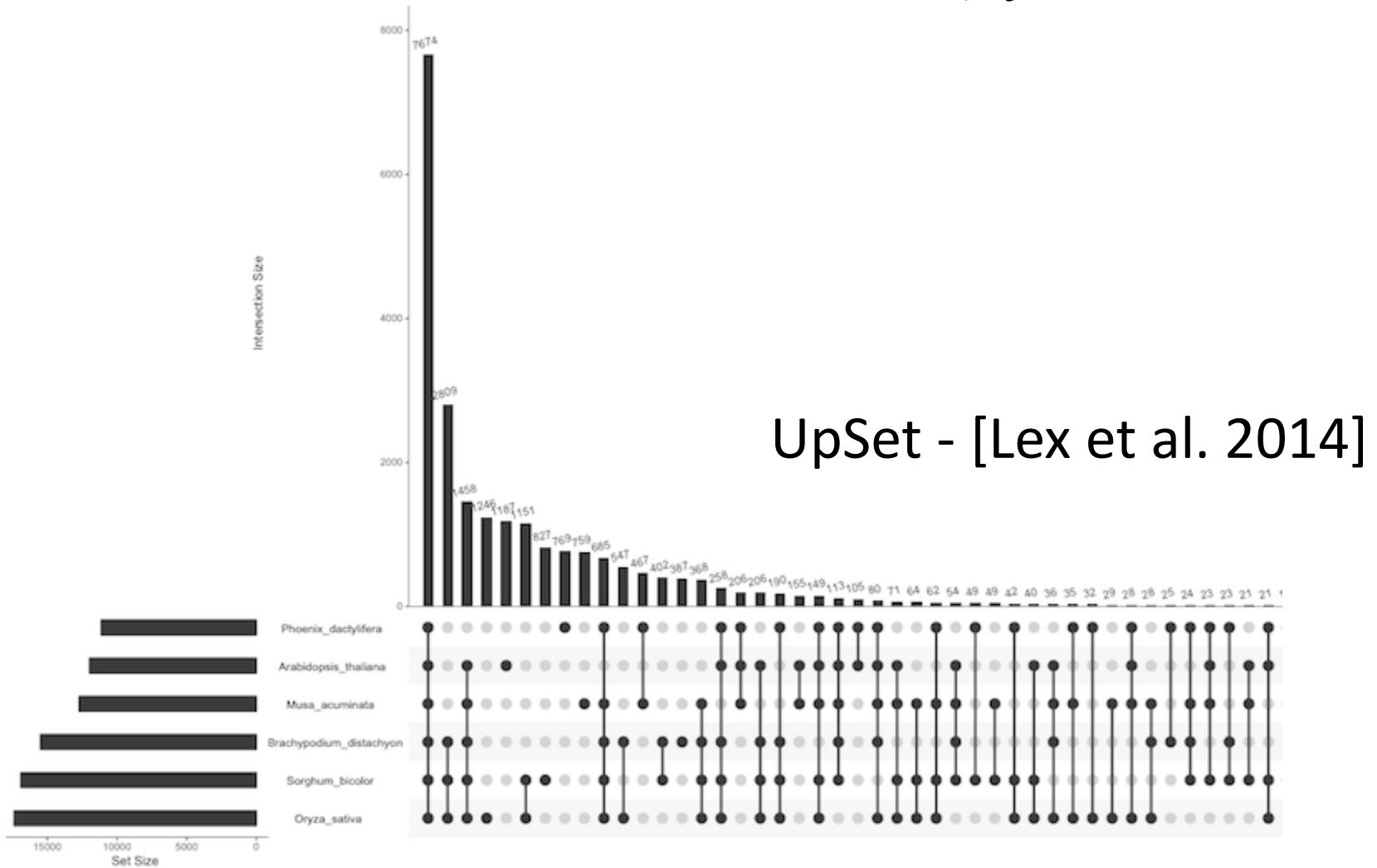
Data Level / Granularity

We assume a simple hierarchical data model:

- $\{attrib_1, ..., attrib_i\}$ = data item

- $\{item_1, ..., item_j\}$ = data cluster

- $\{cluster_1, ..., cluster_k\}$ = data set

- $\{set_1, ..., set_l\}$ = data landscape

Overview -> Detail: We often need to compare at multiple levels!
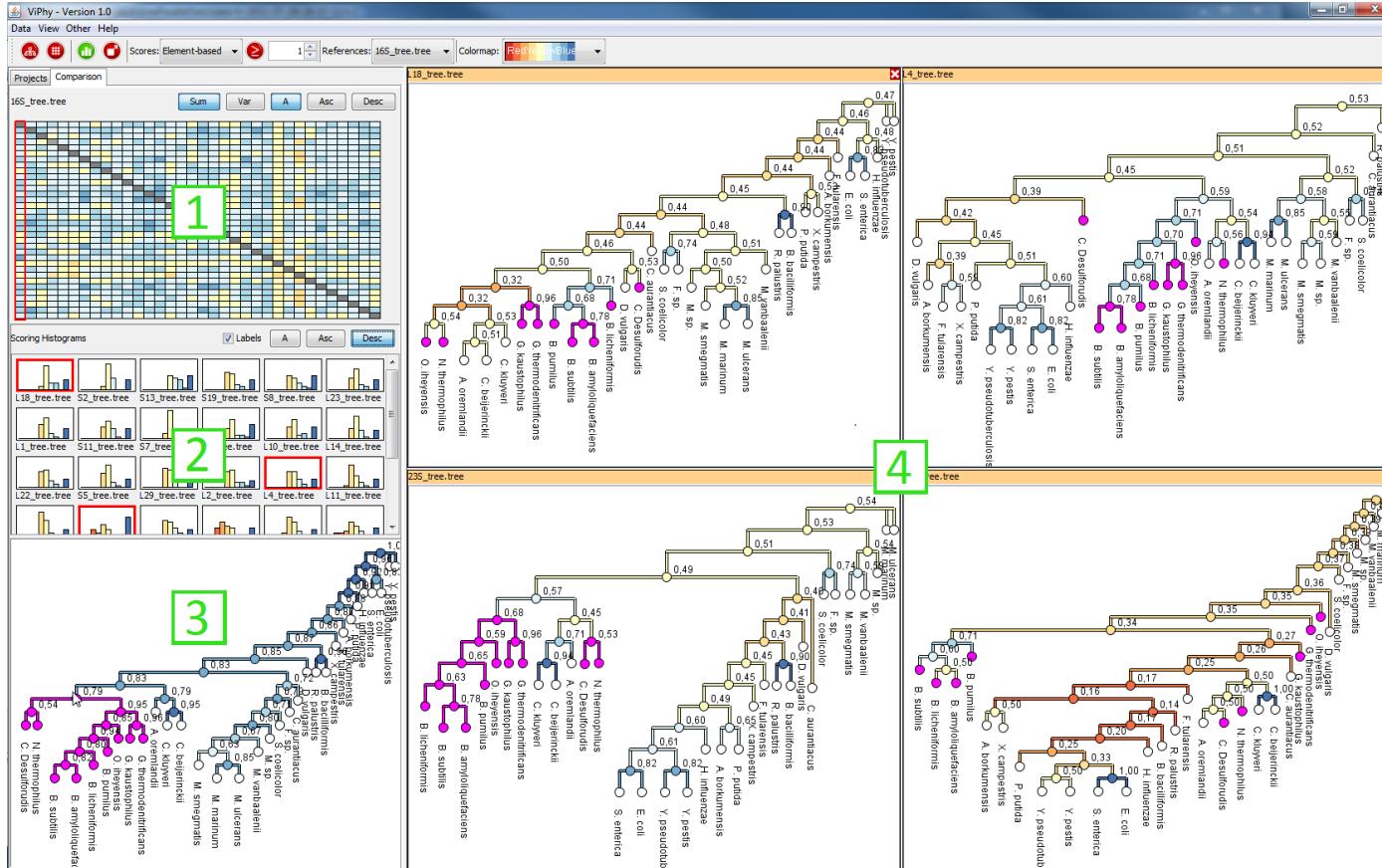

[Bremm et al. 2011]

Comparison Tasks Deconstructed

- WHY to compare? – Goal (objective/aim/purpose)
  - Exploration
  - Confirmation
  - Presentation

- HOW to compare? – Means (action, task)
  - Search / Match
  - Sort / Order
  - Group / Relate

- WHAT to compare? – Data characteristics (data aspect)
  - Data type
  - Data level / granularity
  - Data cardinality

Modeled after [Schulz et al. 2013]

Data Cardinality -> Comparison Type
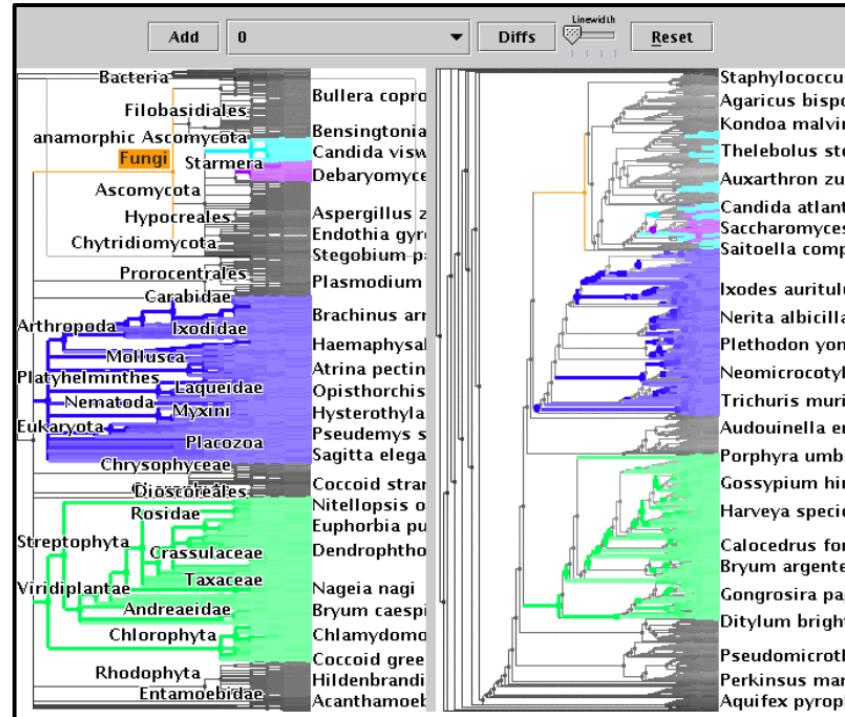
- ## 1:1 Comparison

TreeJuxtaposer [Munzner et al. 2003]

# Data Cardinality -> Comparison Type

- 1:1 Comparison
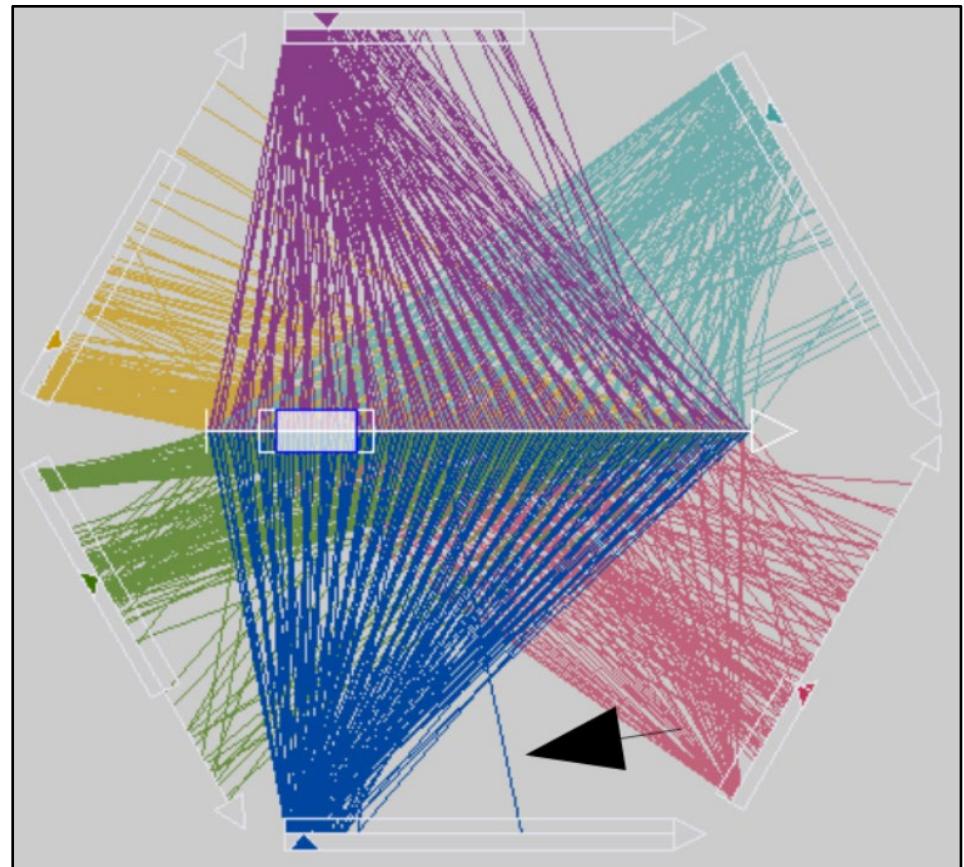
- 1:many
  Comparison

VisAxes [Tominski et al. 2004]

Data Cardinality -> Comparison Type

- 1:1 Comparison

- 1:many Comparison

- many:many Comparison
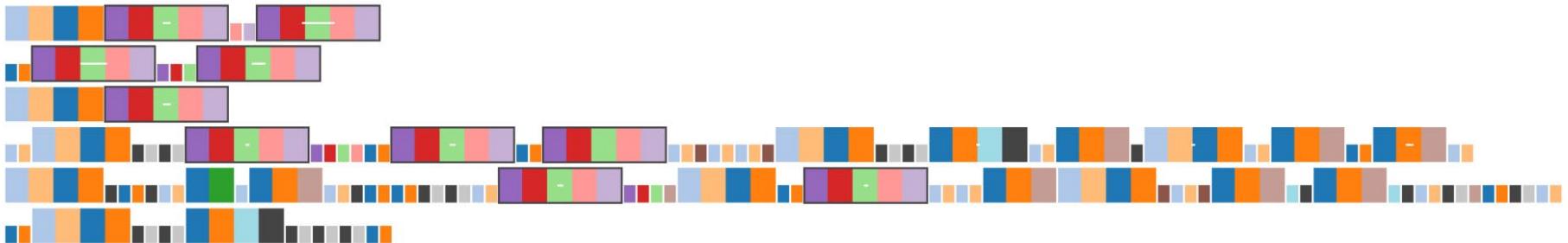
Action Sequences [Nguyen et al. 2018]