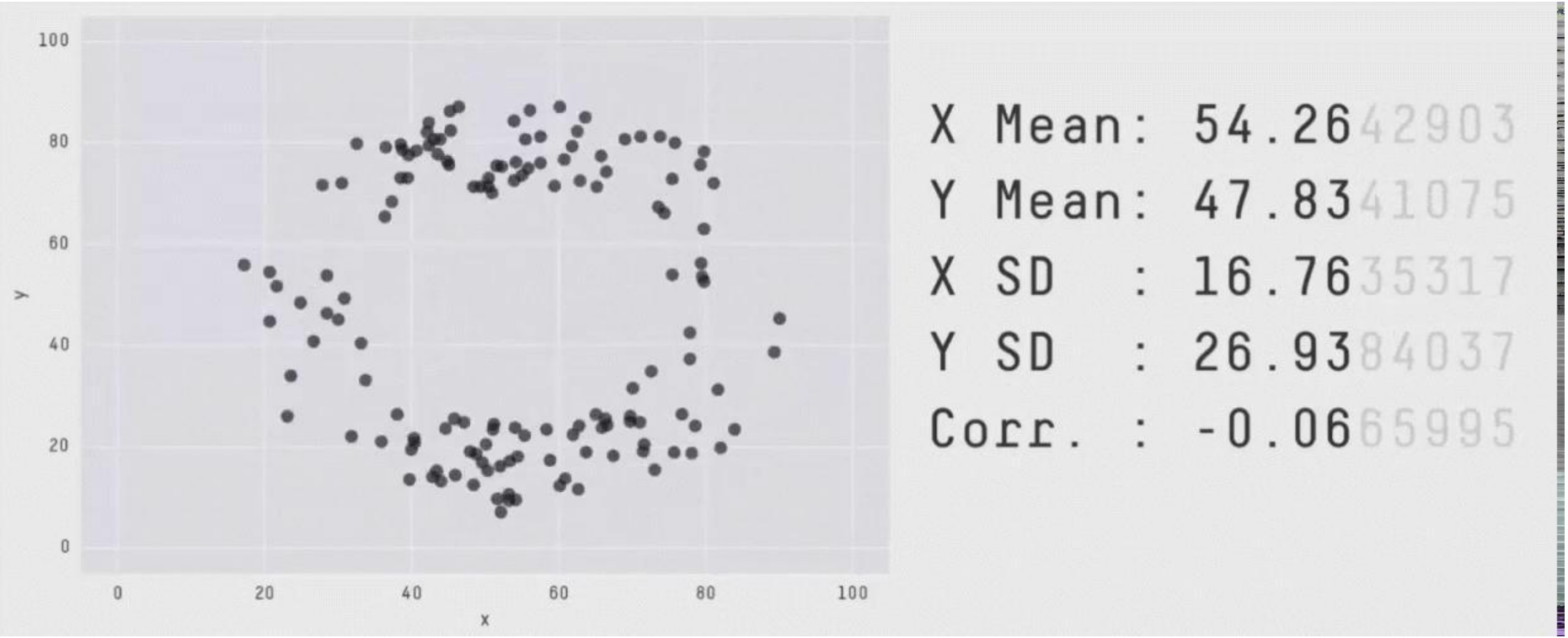
Data visualization

An introduction

Chen He
chen.he@helsinki.fi

Why do we need visualization?



Matejka, J. and Fitzmaurice, G., 2017, May. Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 1290-1294). ACM.

What is **data visualization**?

The use of computer-supported, interactive, visual representations of data to **amplify cognition**.

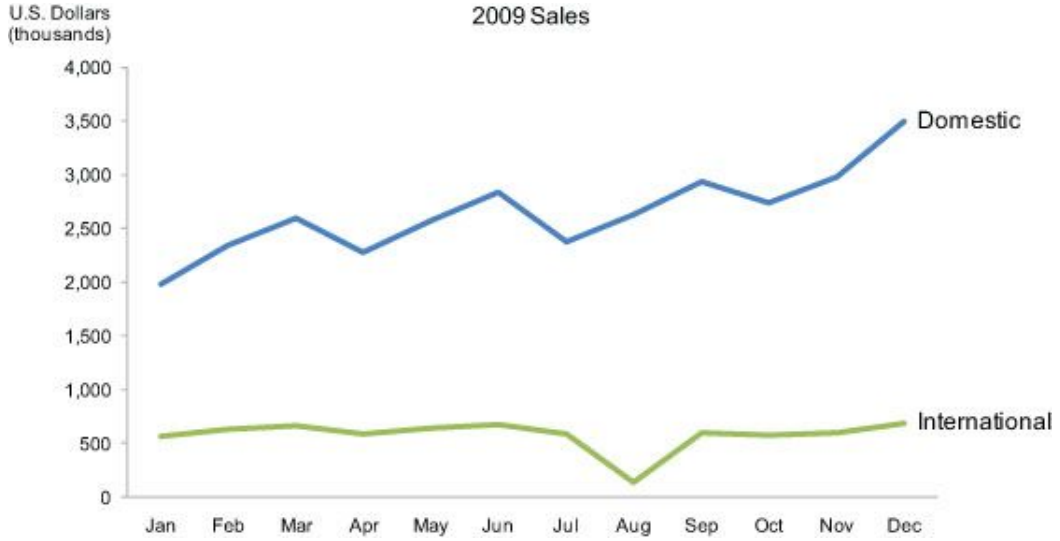
[Card et al., 1999, Readings in Information Visualization]

https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception

Use perception to amplify cognition

2009 Sales (thousands of U.S. $)

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493	31,783
International	574	636	673	593	644	679	593	139	599	583	602	690	7,005
Total	2,557	2,979	3,266	2,876	3,218	3,517	2,975	2,773	3,537	3,322	3,585	4,183	38,788



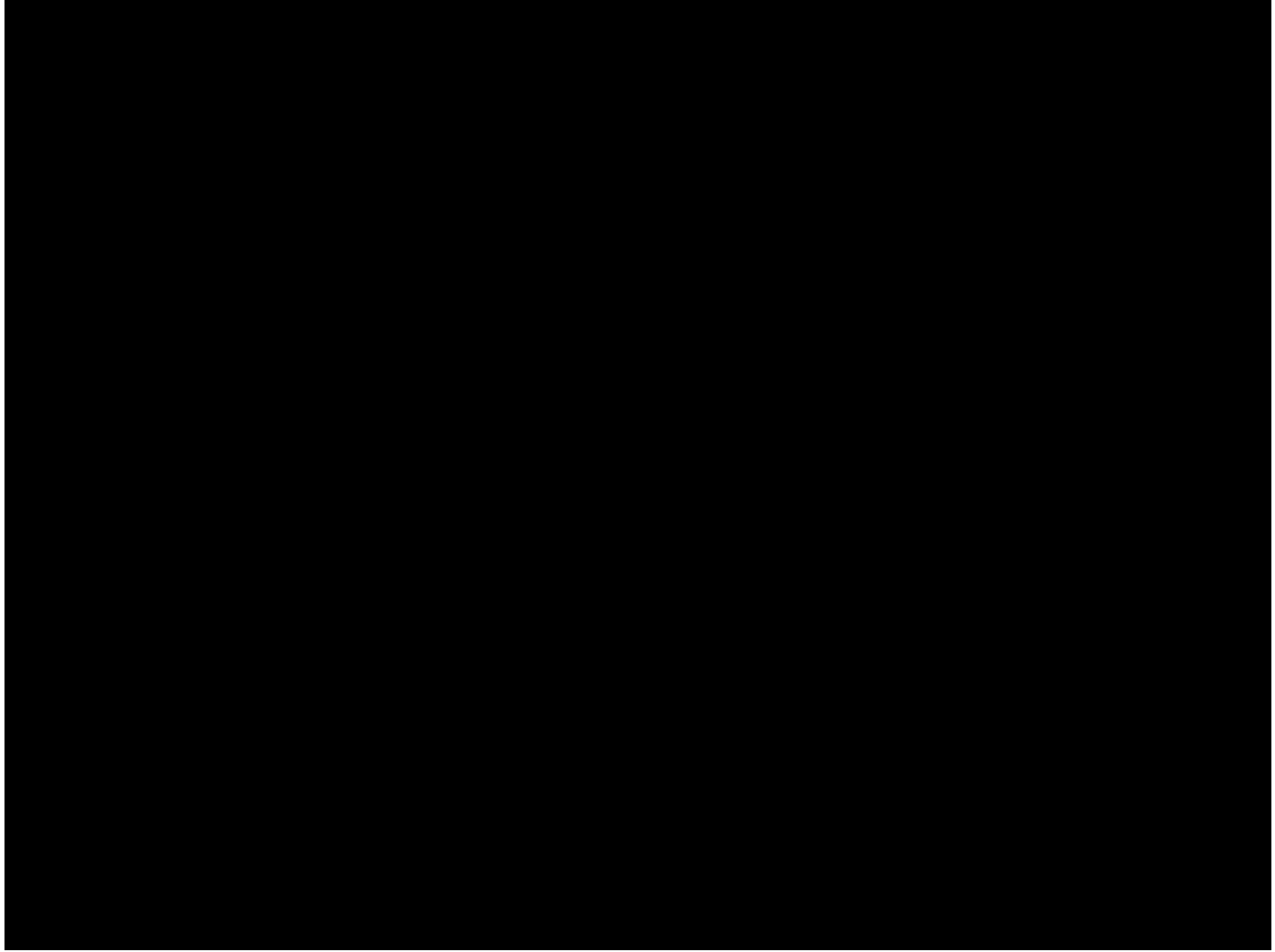
Purposes of visualization

To help

- Make a decision
- Tell a story
- Reasoning
- Discover knowledge
- ……

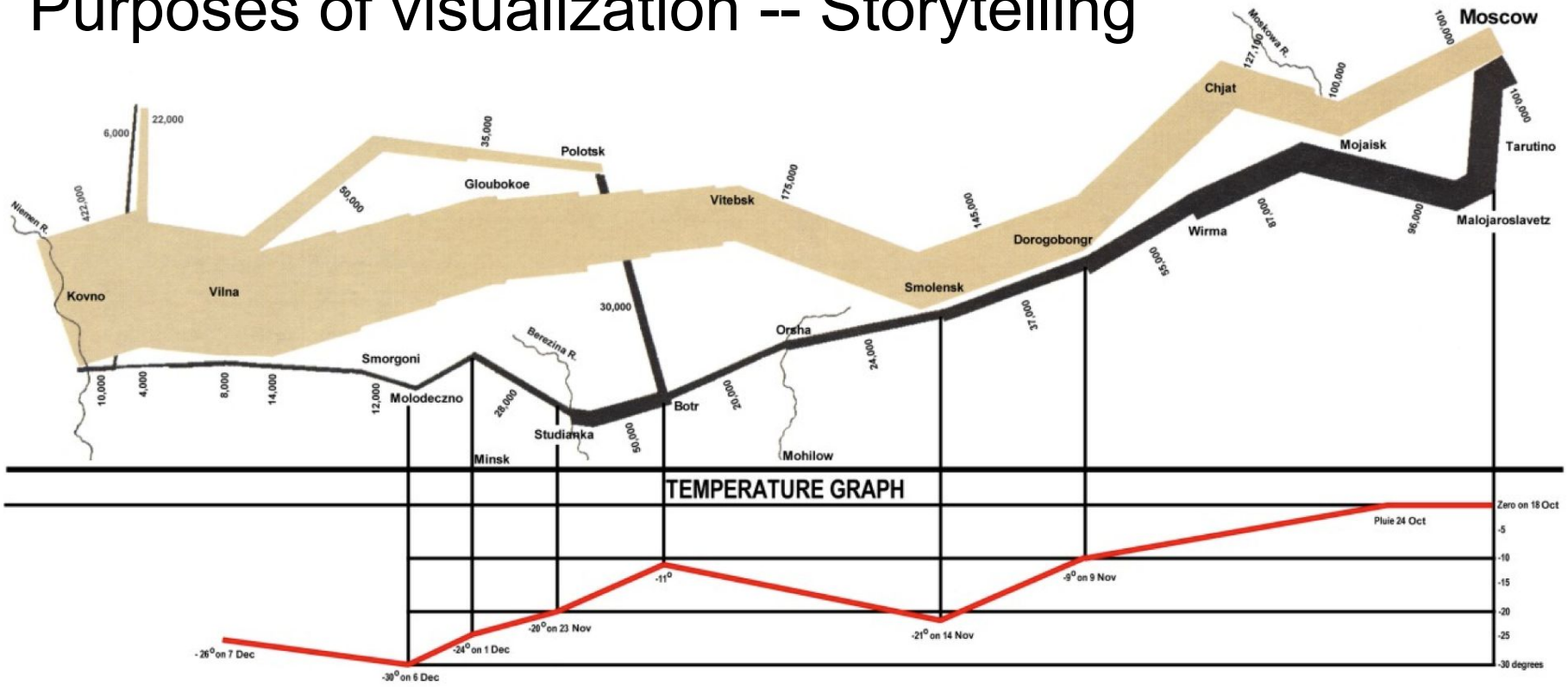
Bostandjiev, S., O'Donovan, J. and Höllerer, T., 2012, September. TasteWeights: a visual interactive hybrid recommender system. In Proceedings of the sixth ACM conference on Recommender systems (pp. 35-42).

Purposes of visualization -- Decision making



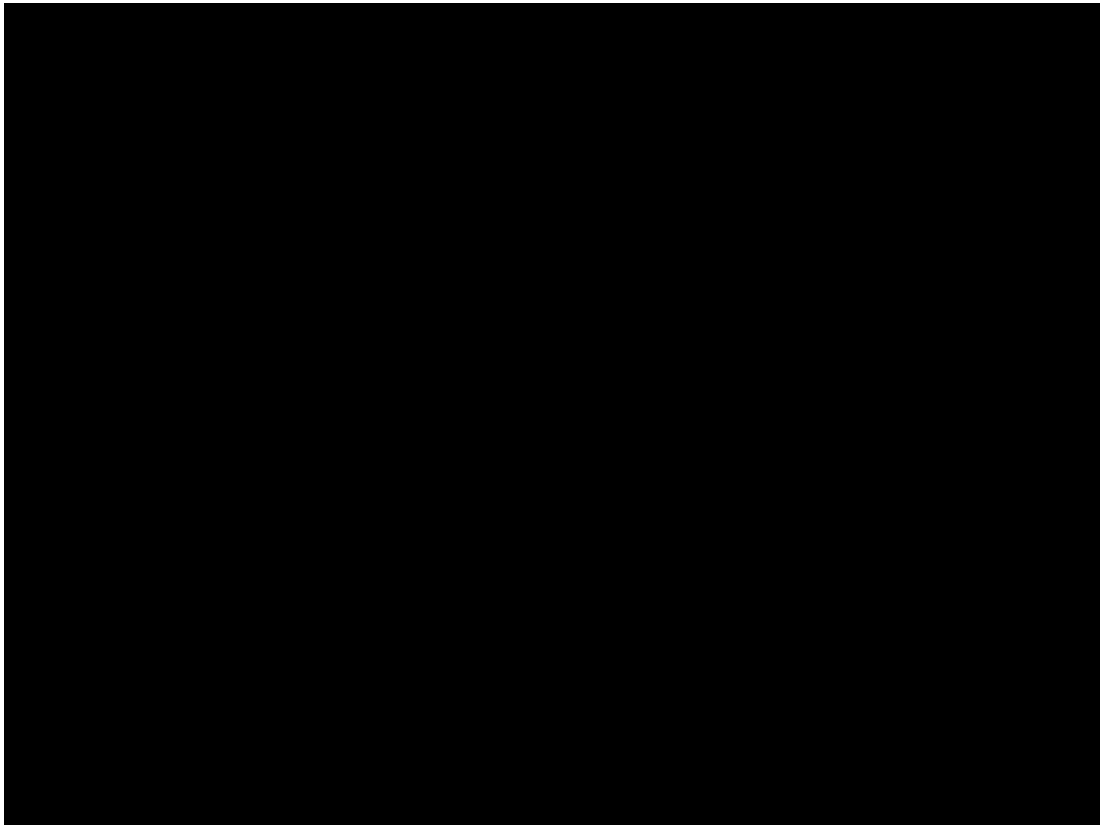
Edward R. Tufte, The Visual Display of Quantitative Information, Graphics Press, Cheshire, CT 1983.

Purposes of visualization -- Storytelling



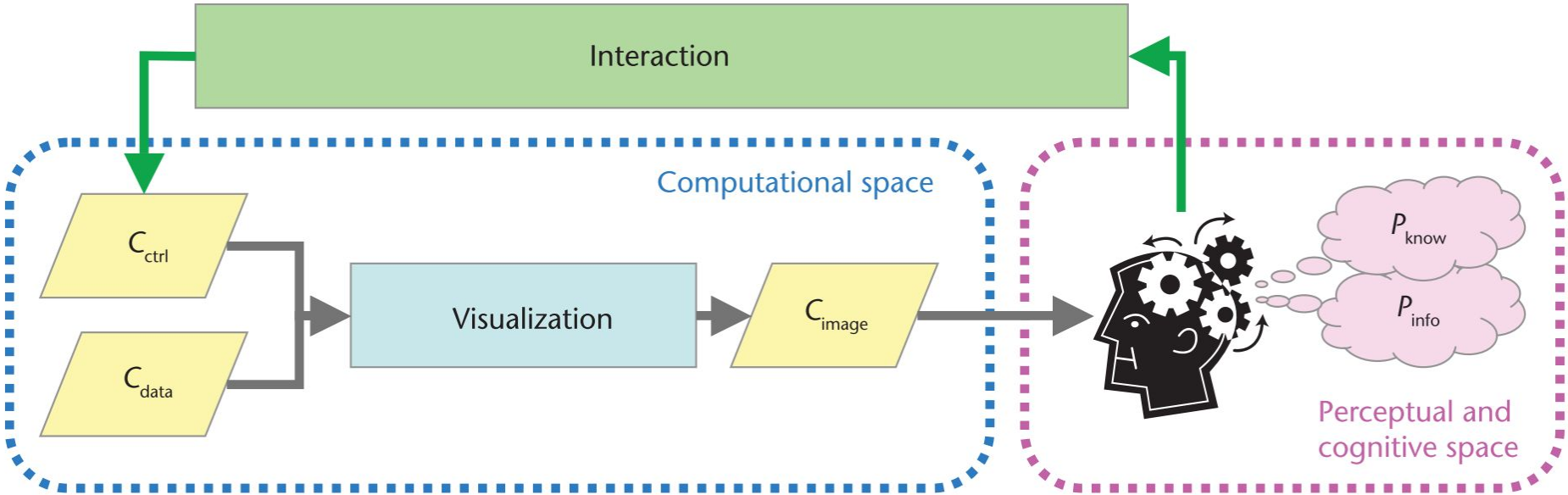
Dork, M., Gruen, D., Williamson, C. and Carpendale, S., 2010. A visual backchannel for large-scale events. IEEE transactions on visualization and computer graphics, 16(6), pp.1129-1138.

Purposes of visualization -- Exploration & discovery



Chen Min et al. Data, information, and knowledge in visualization. CG&A, 2008.

A Visualization Process



Example techniques visualizing various types of data

Two-dimensional data

Multi-dimensional data

Graph

Hierarchical data

Set-typed data

Example techniques visualizing various types of data

Two-dimensional data

Multi-dimensional data

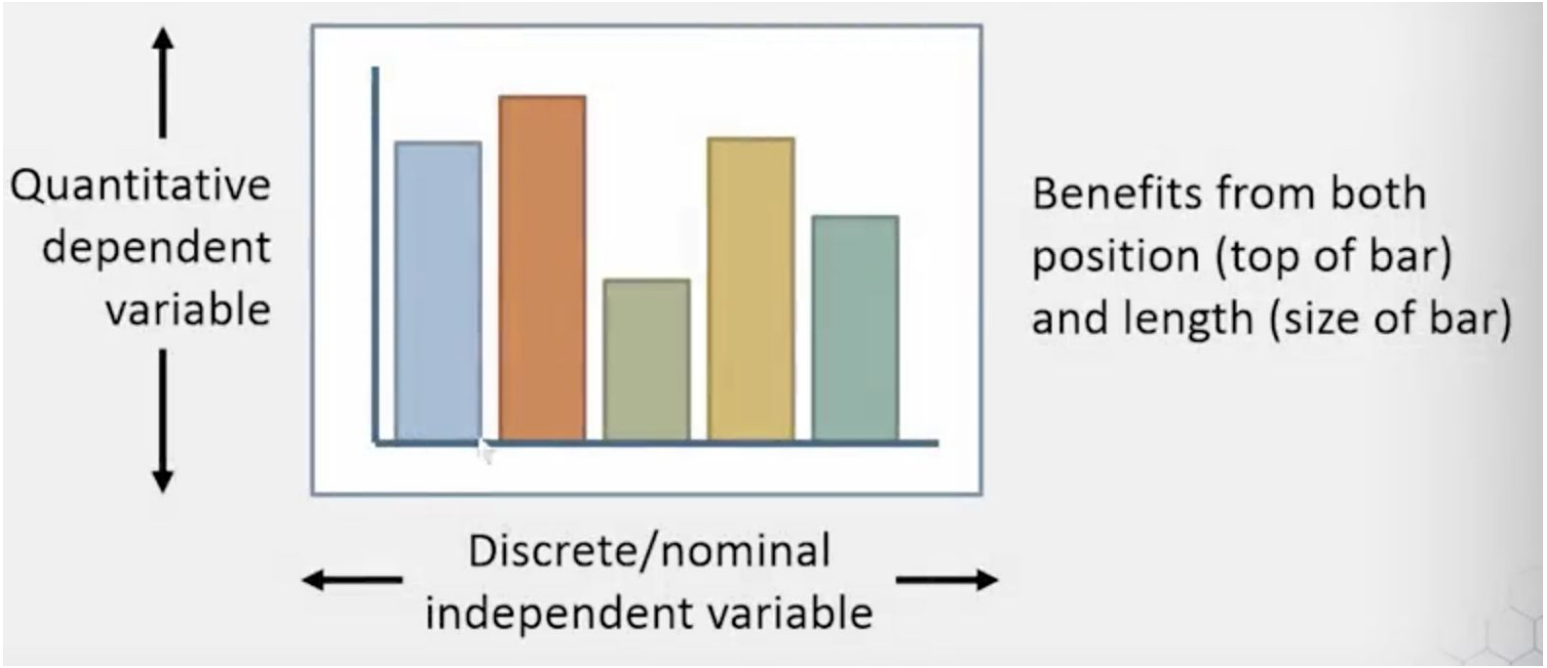
Graph

Hierarchical data

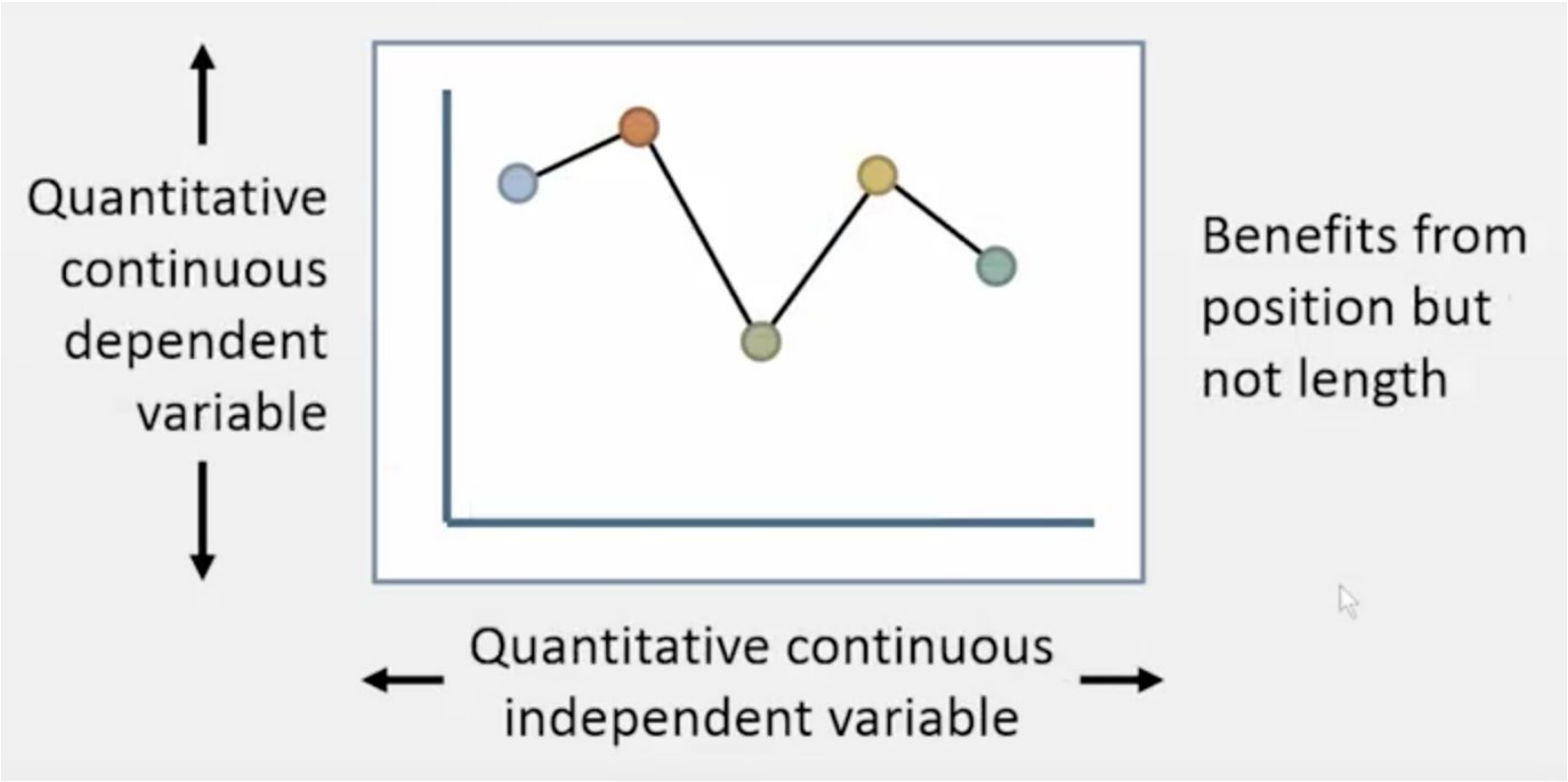
Set-typed data

Following three figures are from https://www.coursera.org/learn/datavisualization

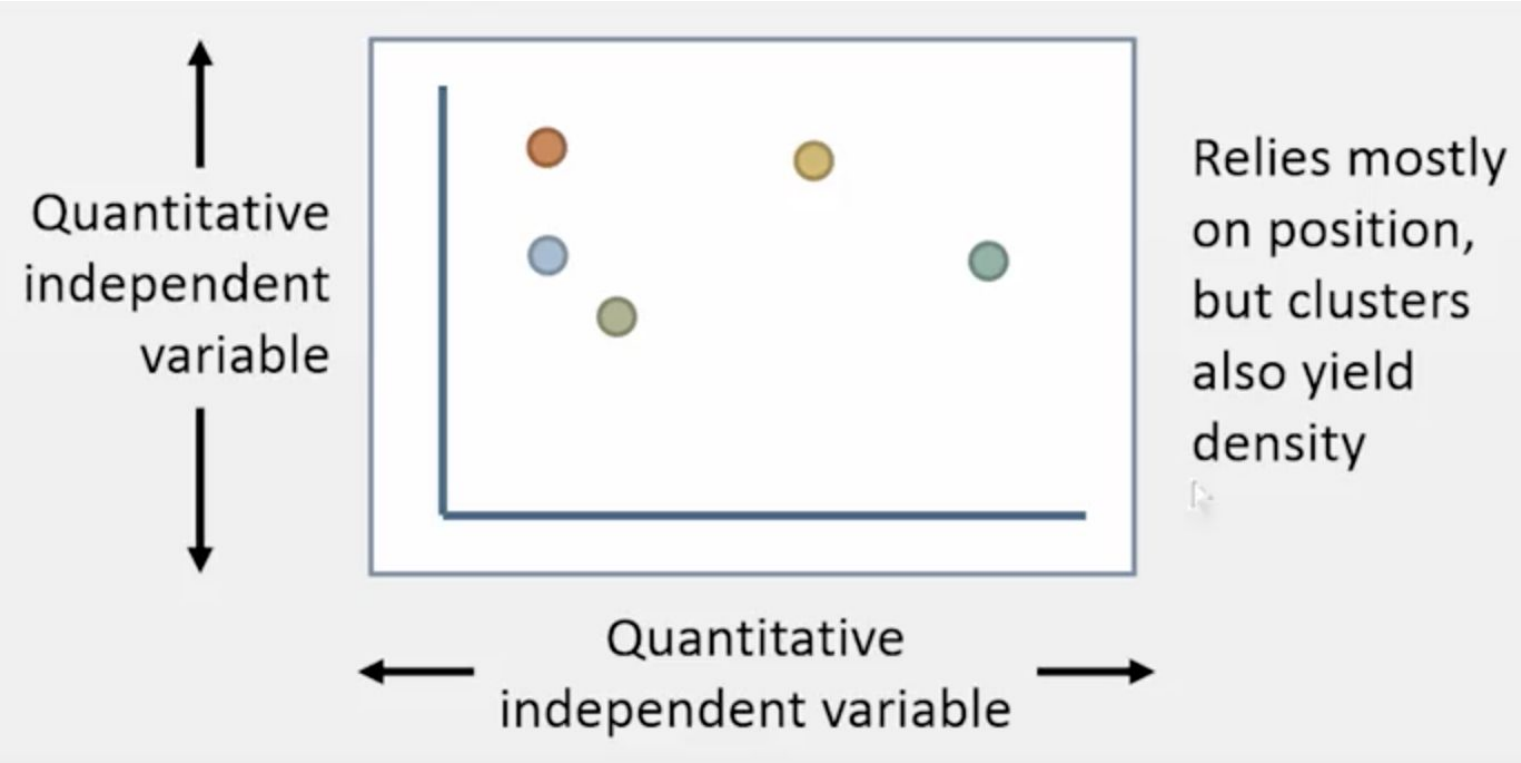
Two-dimensional data -- Bar chart



Two-dimensional data -- Line chart



Two-dimensional data -- Scatterplot



Example techniques visualizing various types of data

Two-dimensional data

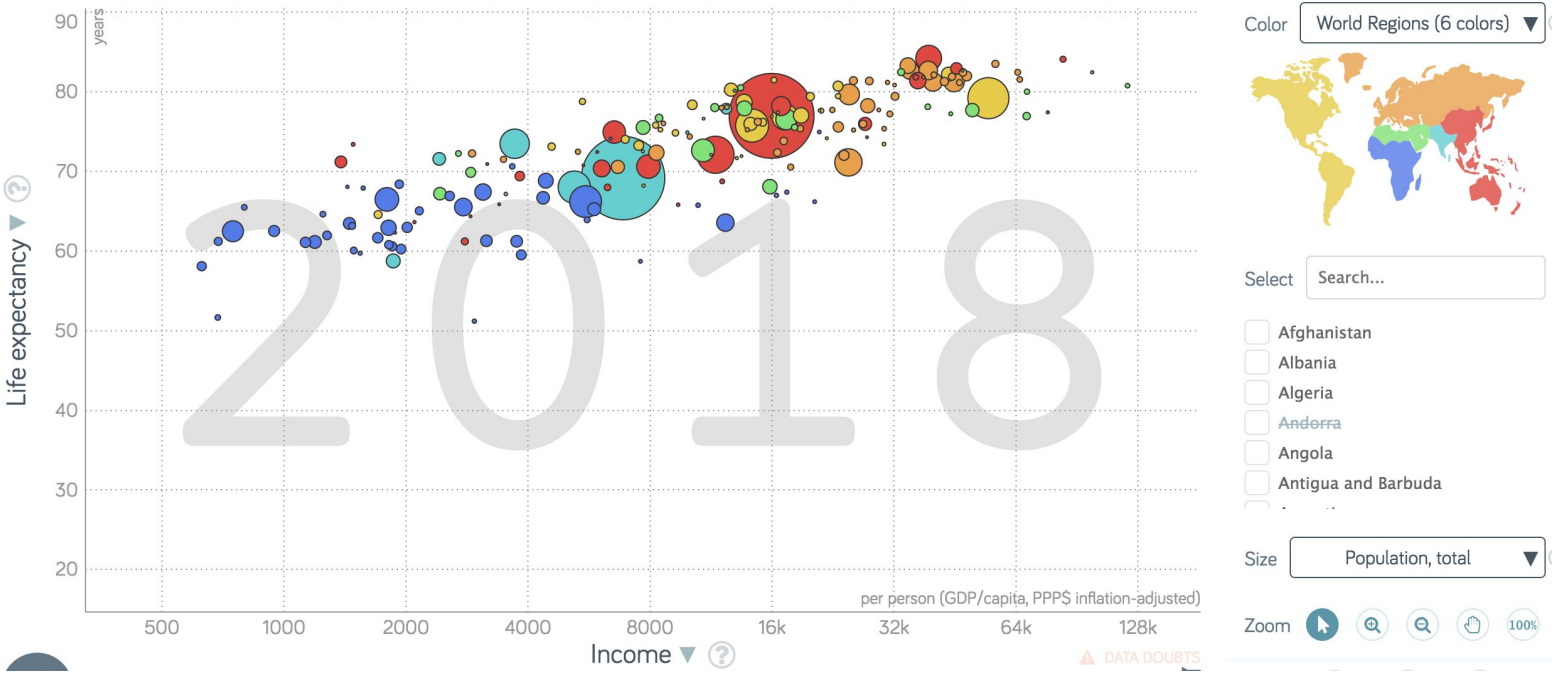
Multi-dimensional data

Graph

Hierarchical data

Set-typed data

Add additional dimensions on top of 2D charts



https://www.gapminder.org/tools/#$chart-type=bubbles

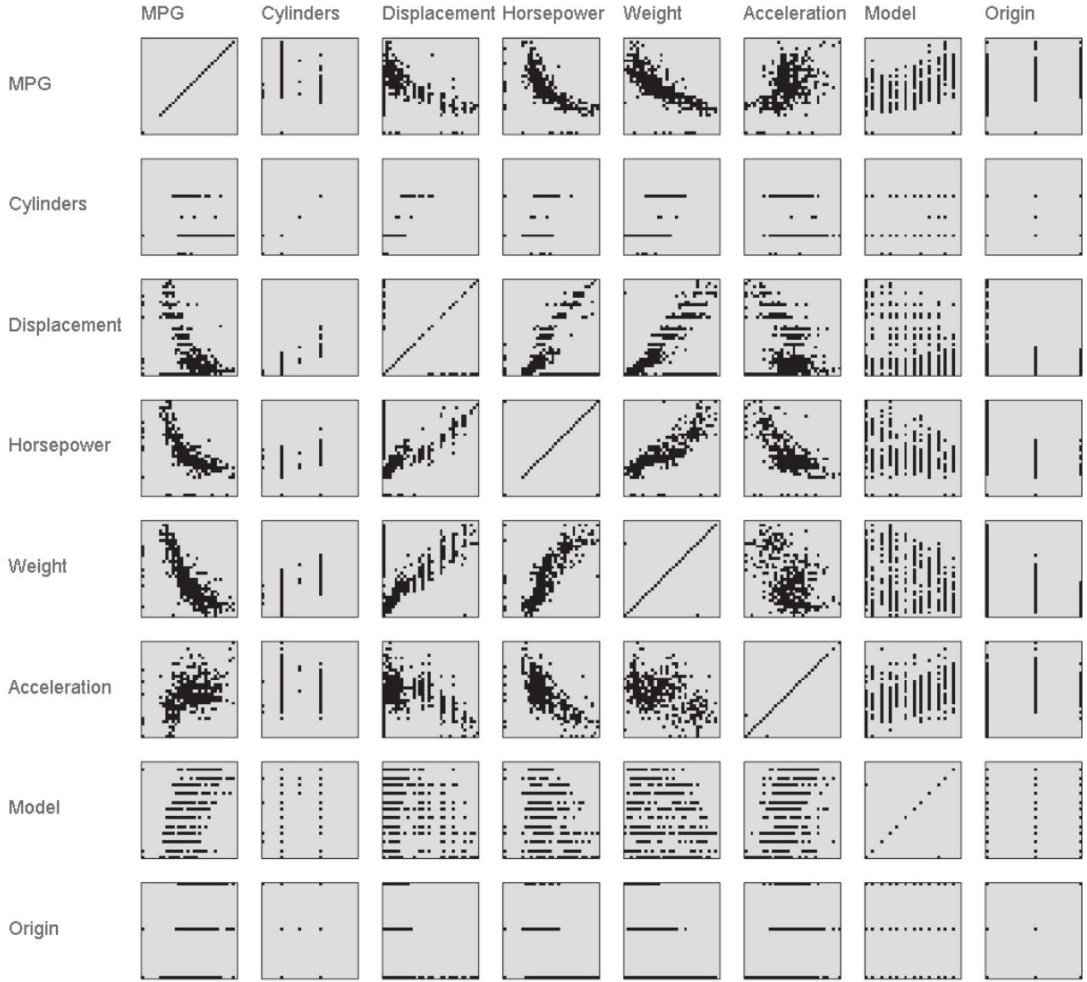
Scatterplot matrix

N variables mean n*n plots.

Diagonal maps the same variable twice.

Each pair is plotted twice, once on each side of the diagonal.

Allows convenient sequential browsing of one variable **compared** to all other variables.



Elmqvist et al., 2008. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Vis*, *14*(6), pp.1539-1148.

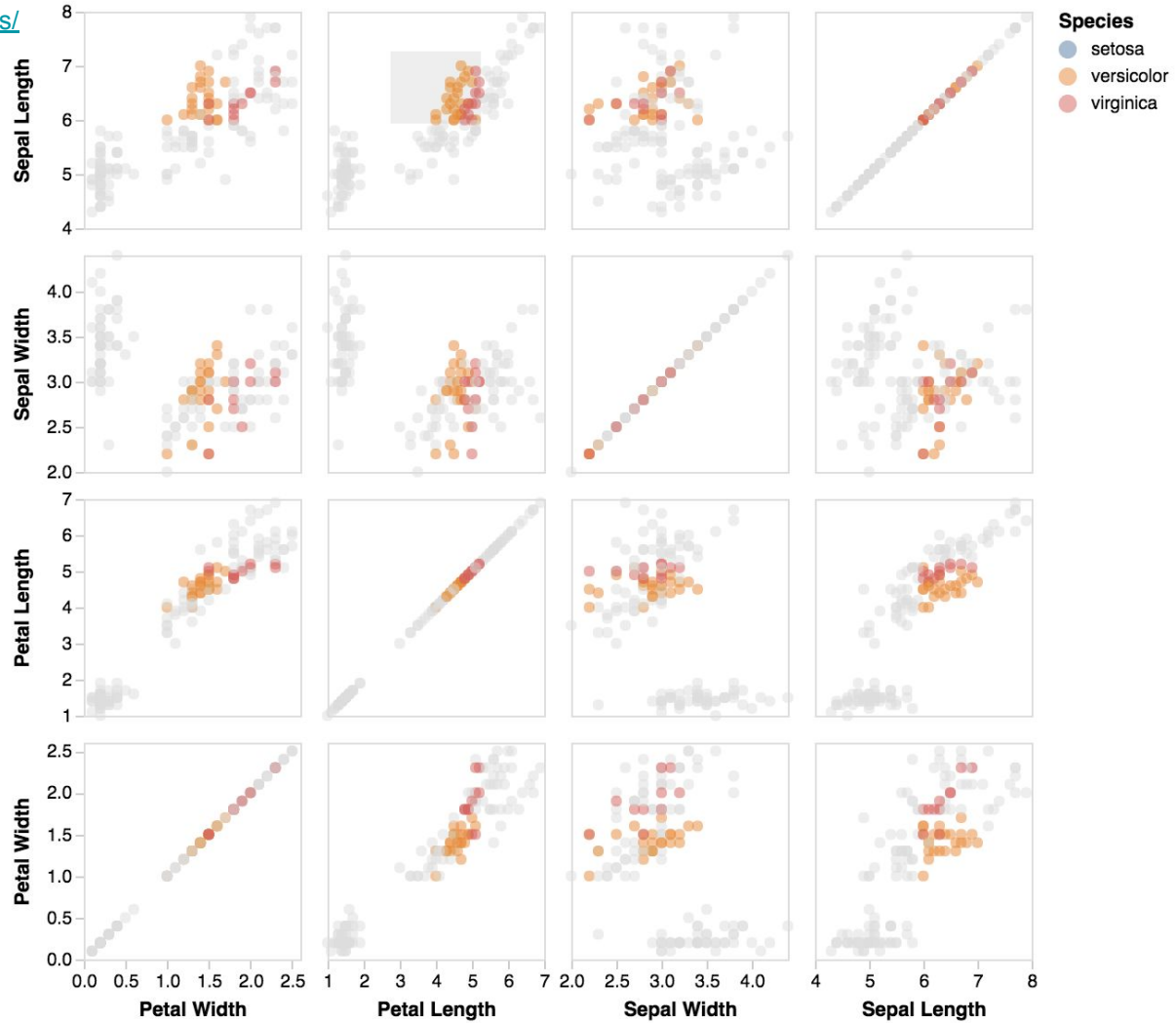
Interacting with multi-dimensional data

Brushing is the process of interactively selecting a subset of data items from a visual representation.

Brushing & linking cause the brush effect (highlighting, etc.) to be applied on those points in the other plots that represent the same data items.

https://vega.github.io/vega/examples/brushing-scatter-plots/

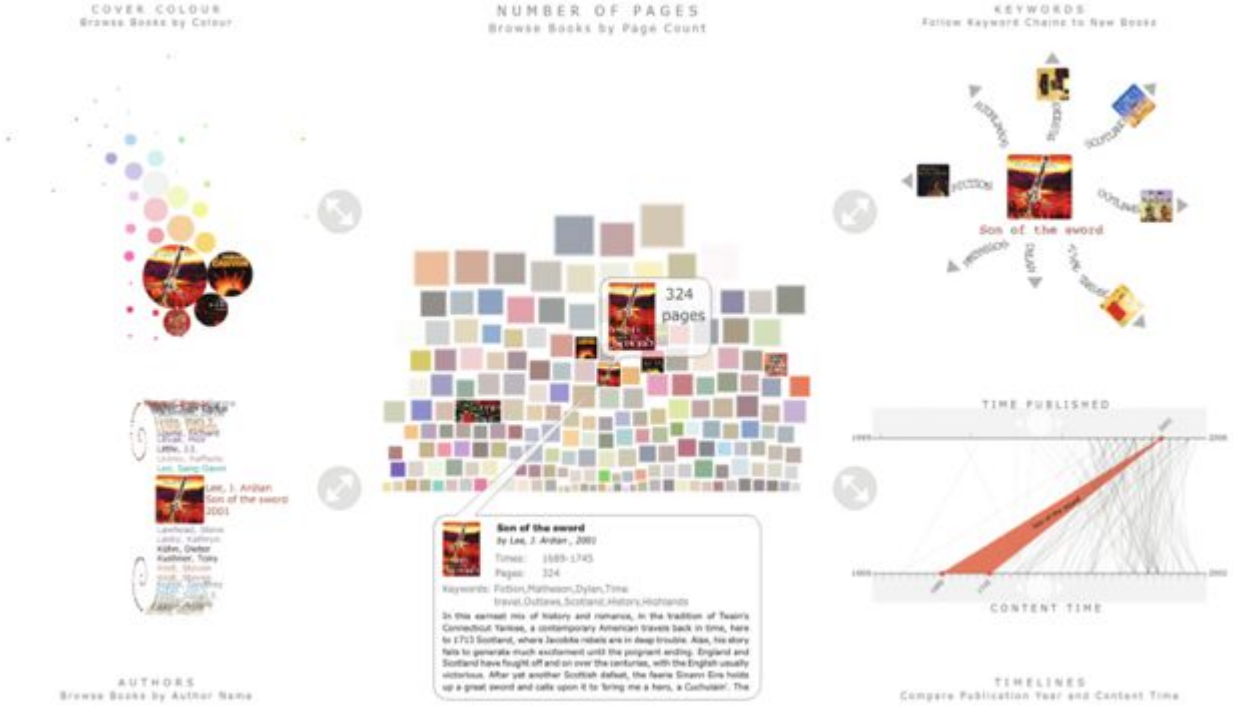
Brushing & linking scatter plots

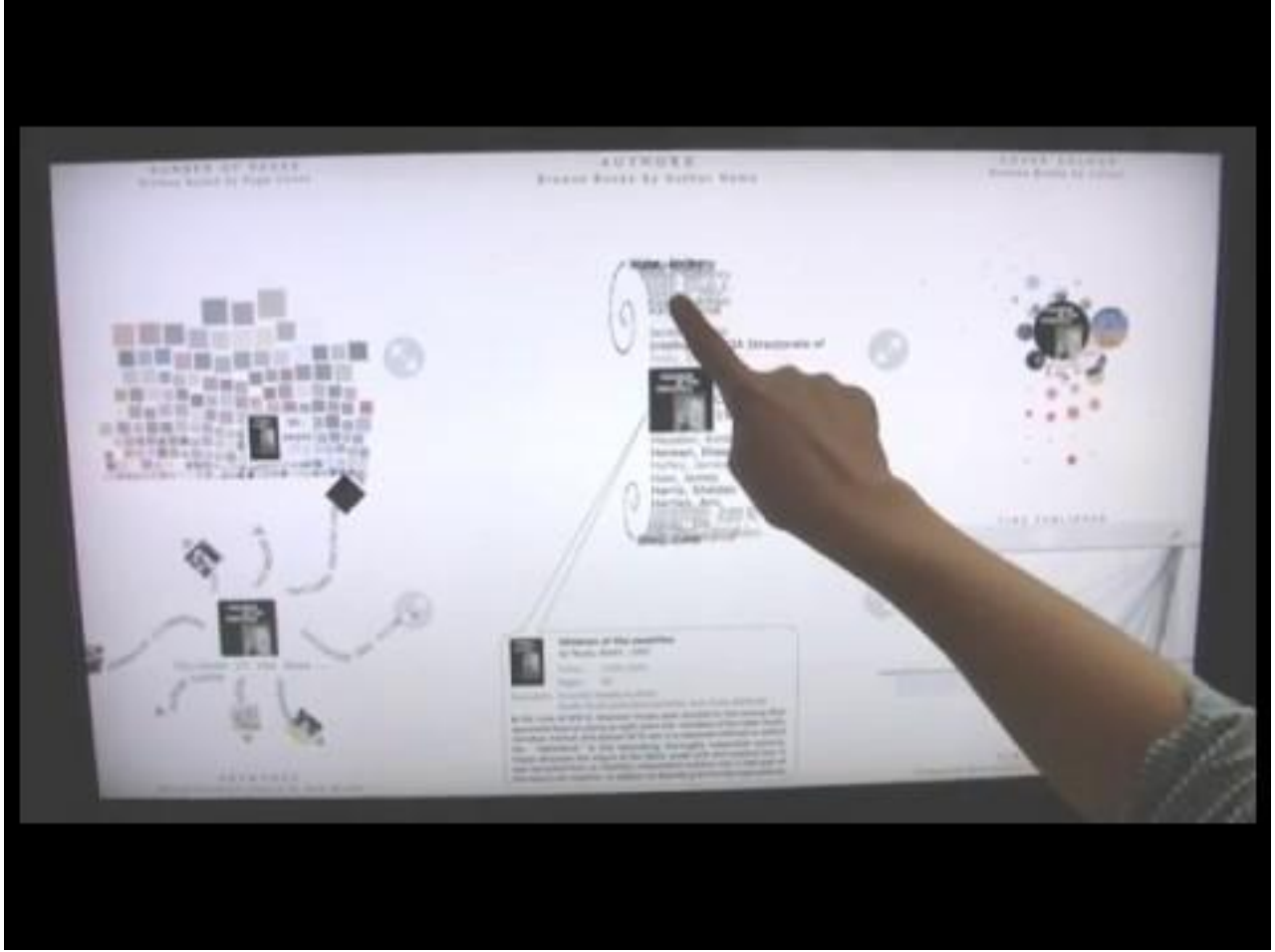


Alice Thudt, Uta Hinrichs and Sheelagh Carpendale. The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization. CHI 2012.

Multi-dimensional data -- Multiple views

Coordinate surprising aspects, like cover color and page numbers, to support serendipitous discovery.





Alice Thudt, Uta Hinrichs and Sheelagh Carpendale. The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization. CHI 2012.

Example techniques visualizing various types of data

Two-dimensional data

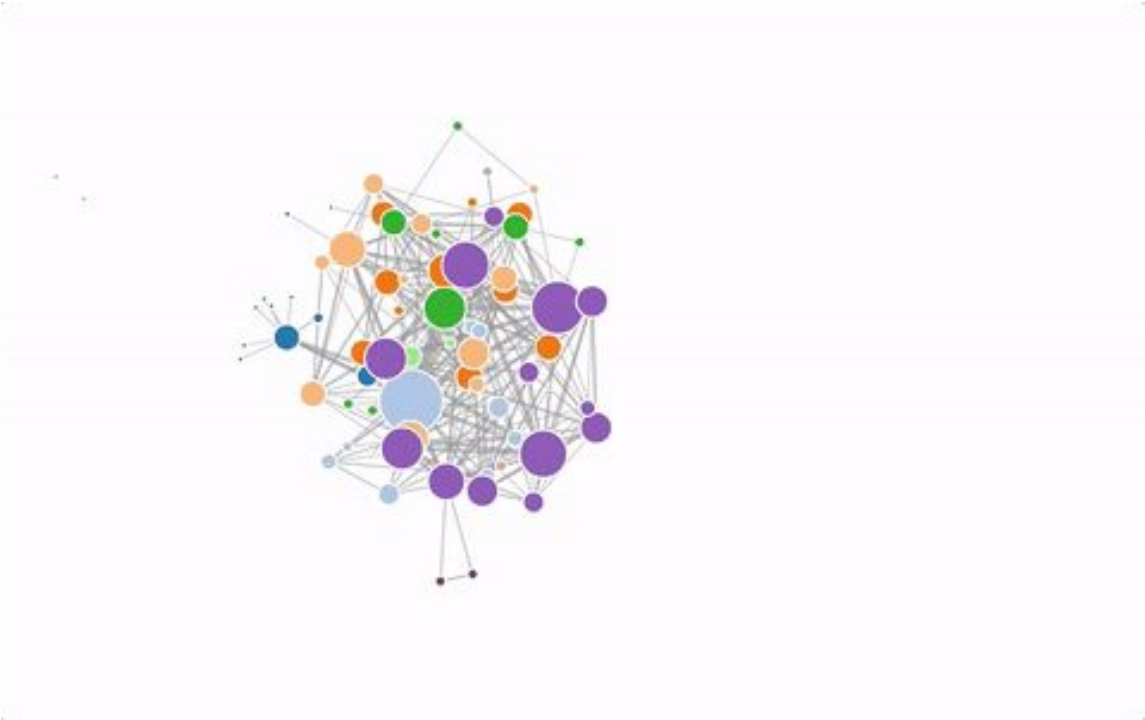
Multi-dimensional data

Graph

Hierarchical data

Set-typed data

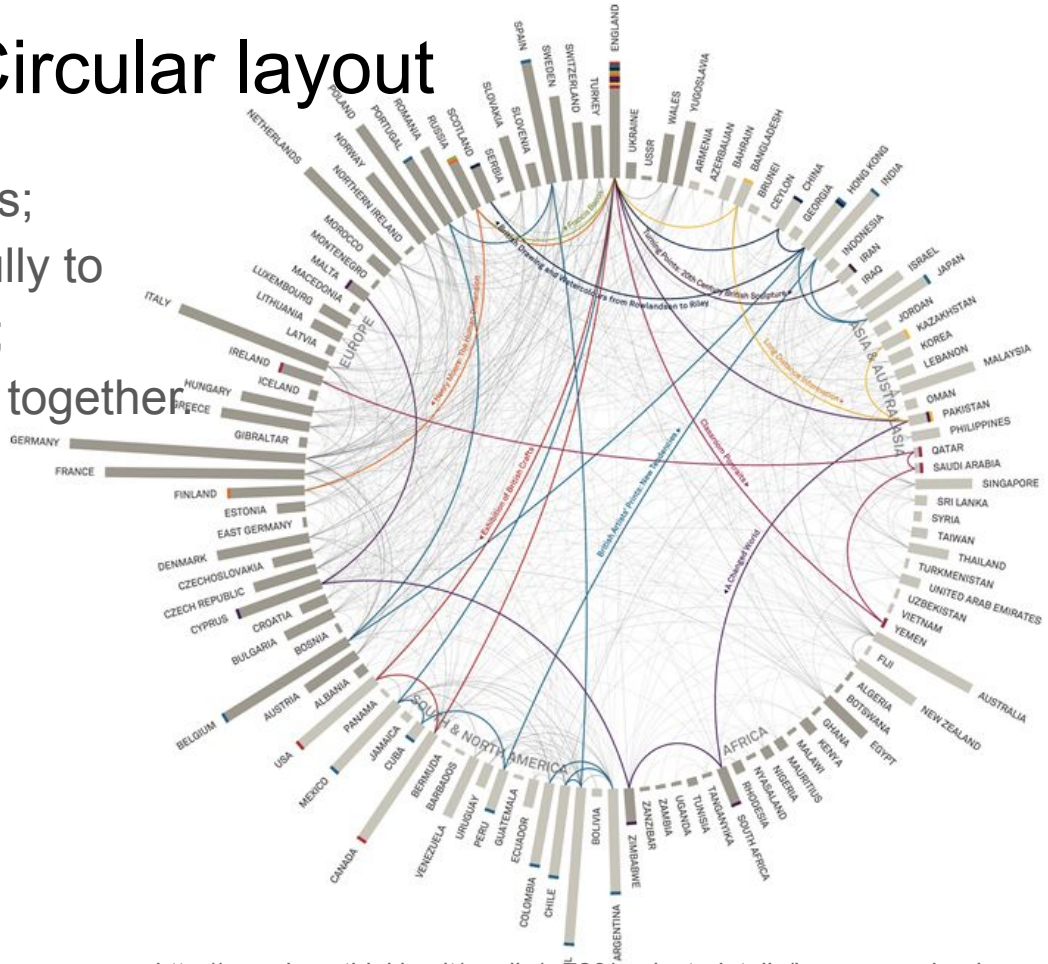
Node-link diagrams -- Force-directed layout



https://bl.ocks.org/heybignick/3faf257bbbbc7743bb72310d03b86ee8

Node-link diagrams -- Circular layout

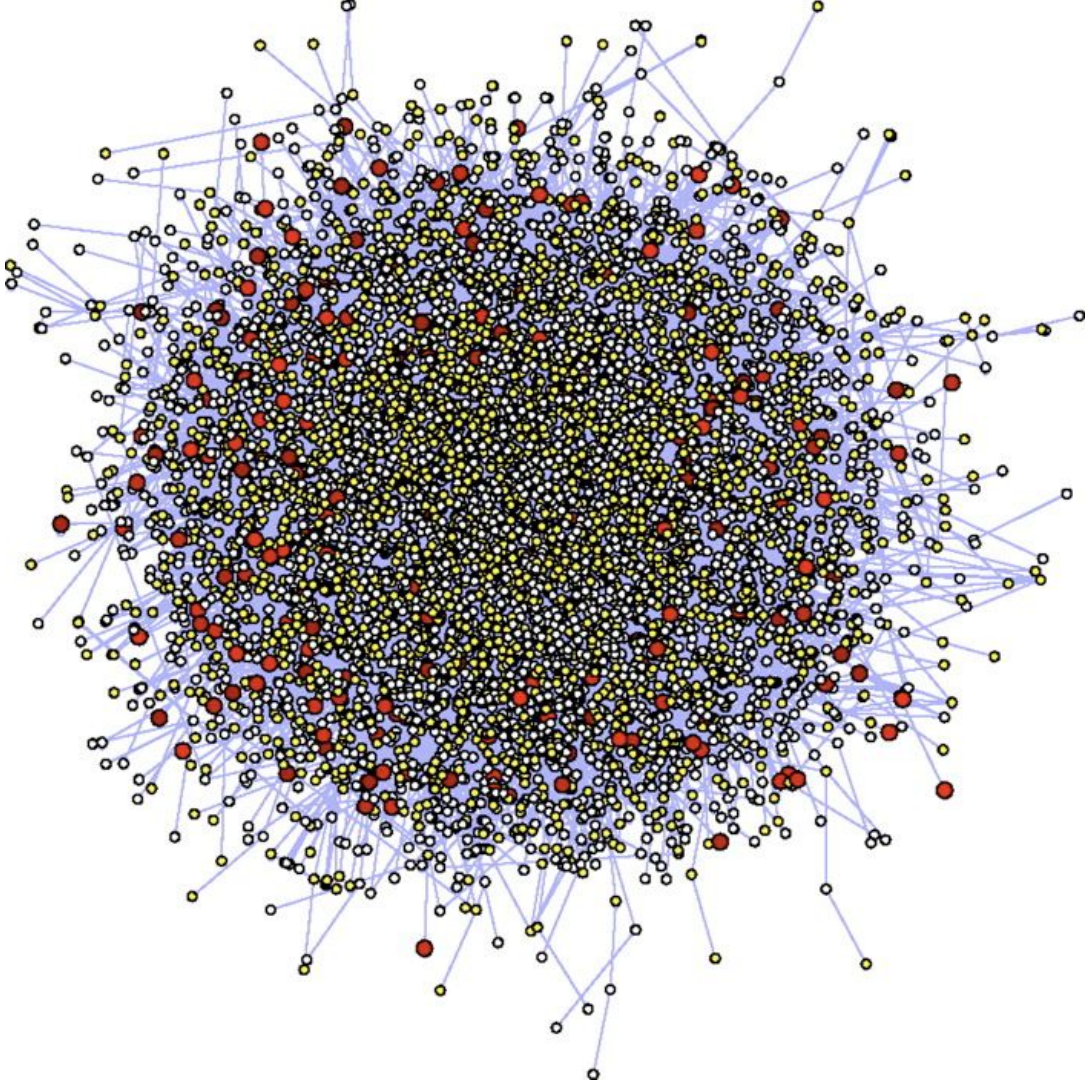
+ can show various node attributes;
- Nodes should be ordered carefully to
  - reduce edge crossings and;
  - place adjacent nodes close together



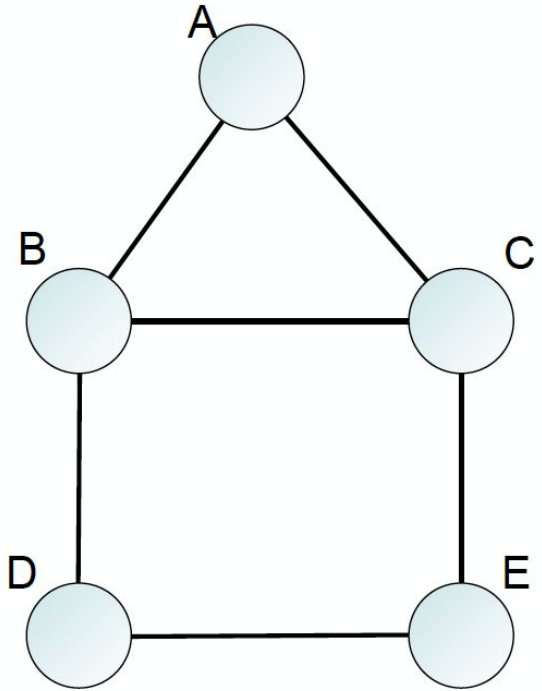
http://www.keepthinking.it/media/w720/project_details/bcva_wc_wheel.png

Node-link diagrams

- \+ Intuitive
- \+ Can show overall structure, clusters, and paths
- \+ Flexible, many variations
- \- Not good for dense graphs
  - Hairball problem

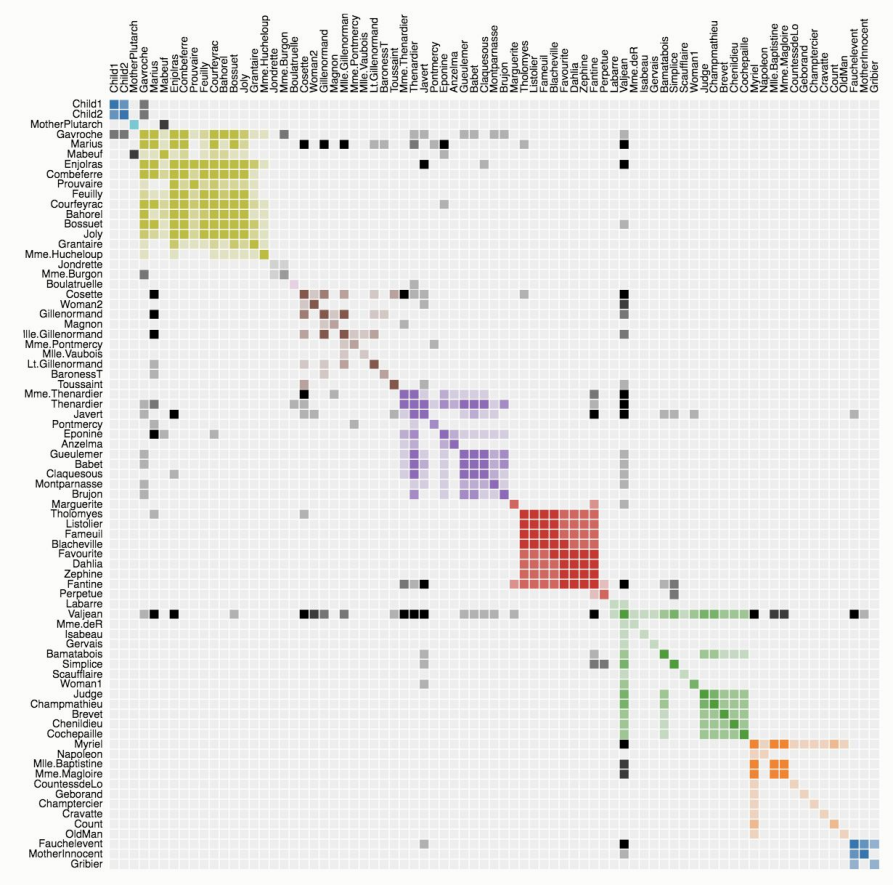


Adjacency matrix



	A	B	C	D	E
A		■	■		
B	■		■	■	
C	■	■			■
D		■			■
E			■	■	

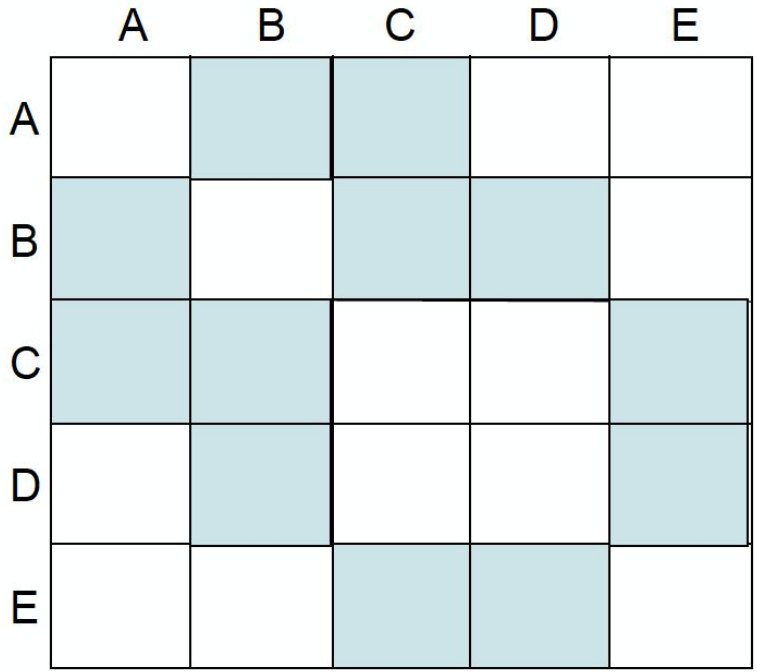
Adjacency matrix



https://bost.ocks.org/mike/miserables/

Adjacency matrix

+ Great for dense graphs
+ Visually scalable
+ Can spot clusters
- Row order affects what you can see
- Abstract visualization
- Path-following is difficult



Example techniques visualizing various types of data

Two-dimensional data

Multi-dimensional data

Graph

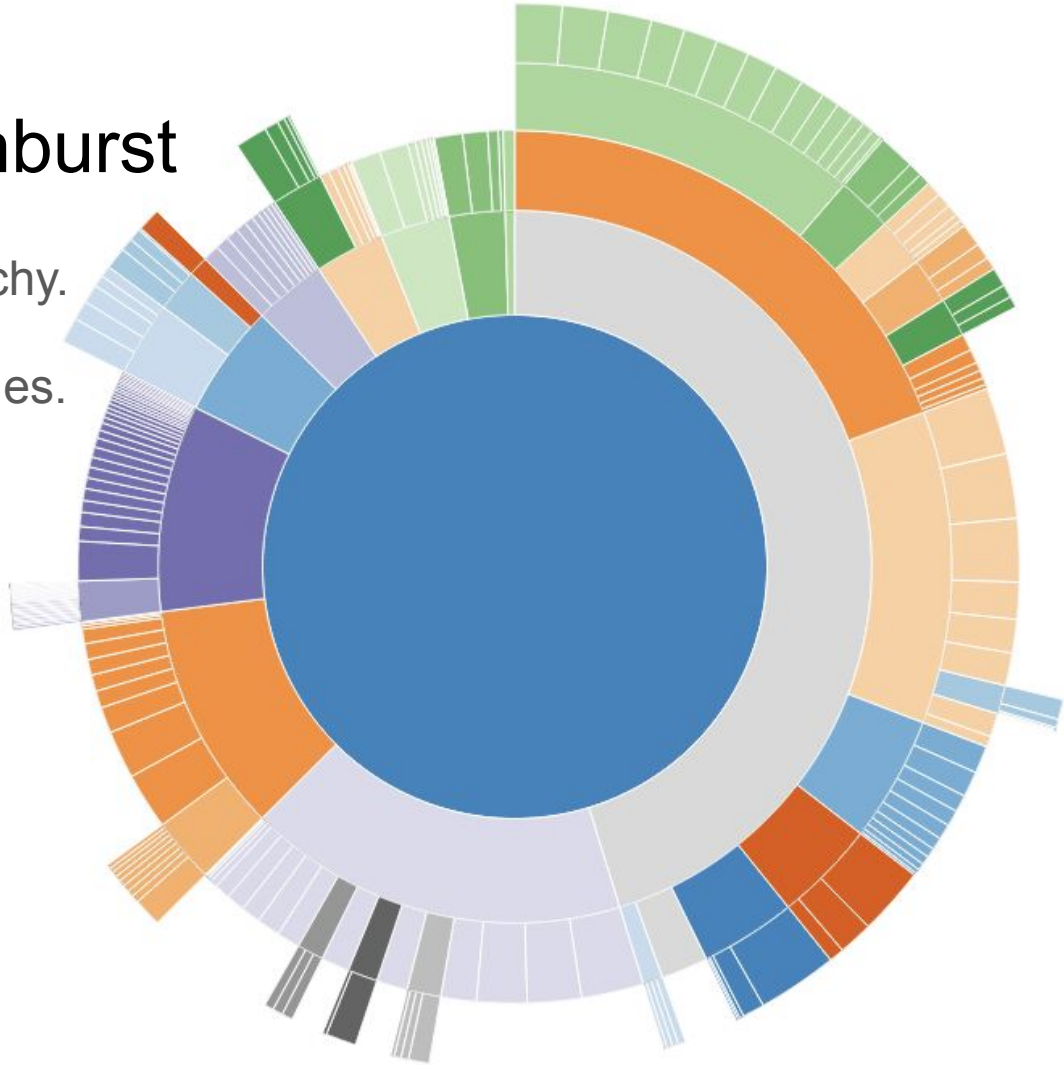
Hierarchical data

Set-typed data

Hierarchical data -- Sunburst

Using adjacency to represent hierarchy.

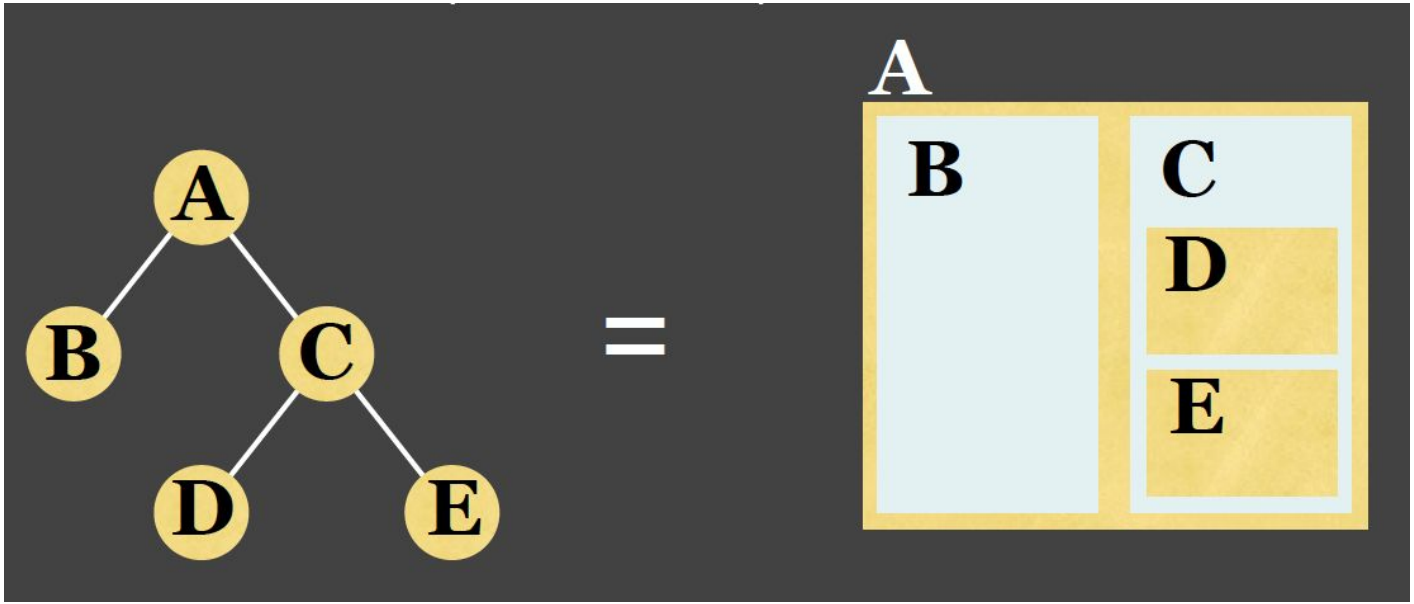
A length encoding for the size of nodes.

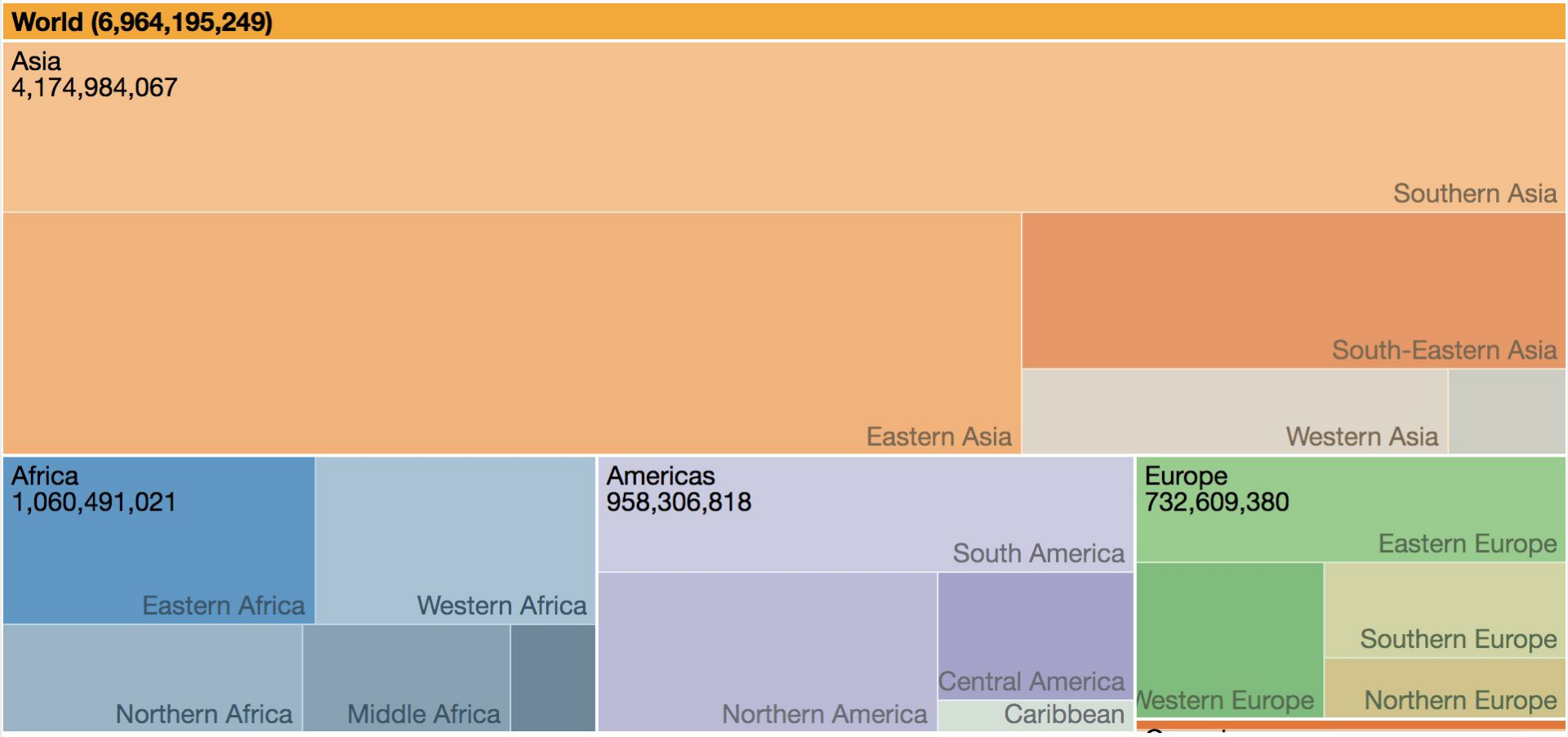


https://bl.ocks.org/kerryrodden/7090426

Hierarchical data -- Treemaps

Using **containment** to represent hierarchy;

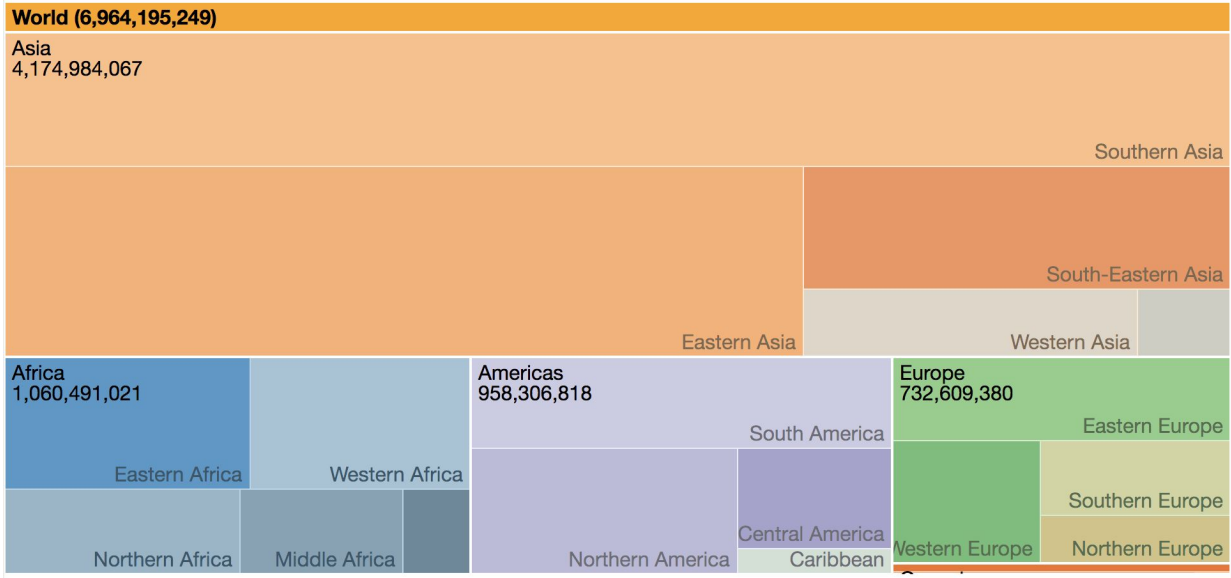




http://bl.ocks.org/ganeshv/6a8e9ada3ab7f2d88022

Hierarchical data -- Treemaps

+ provides single view of entire tree
+ easier to spot small / large node
- difficult to accurately read depth



Example techniques visualizing various types of data

Two-dimensional data

Multi-dimensional data

Graph

Hierarchical data

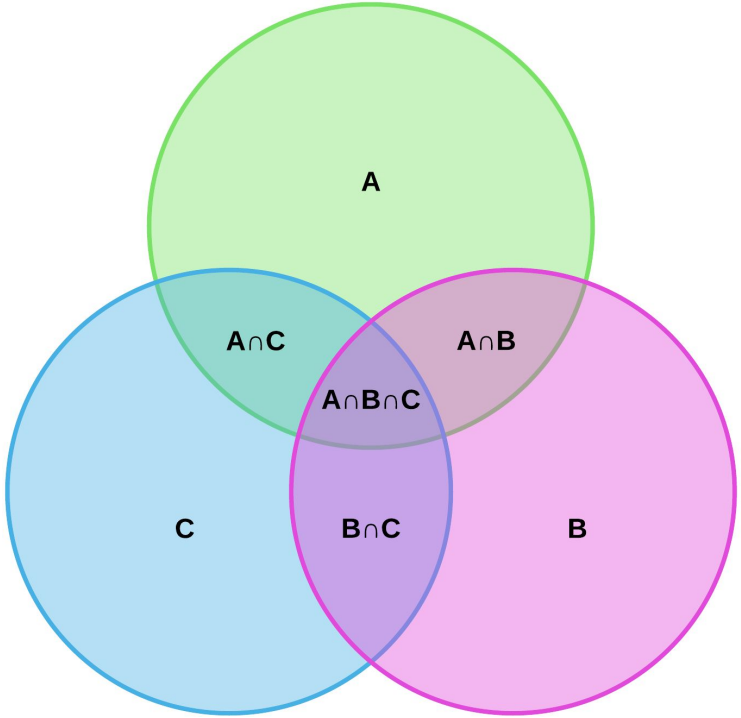
Set-typed data

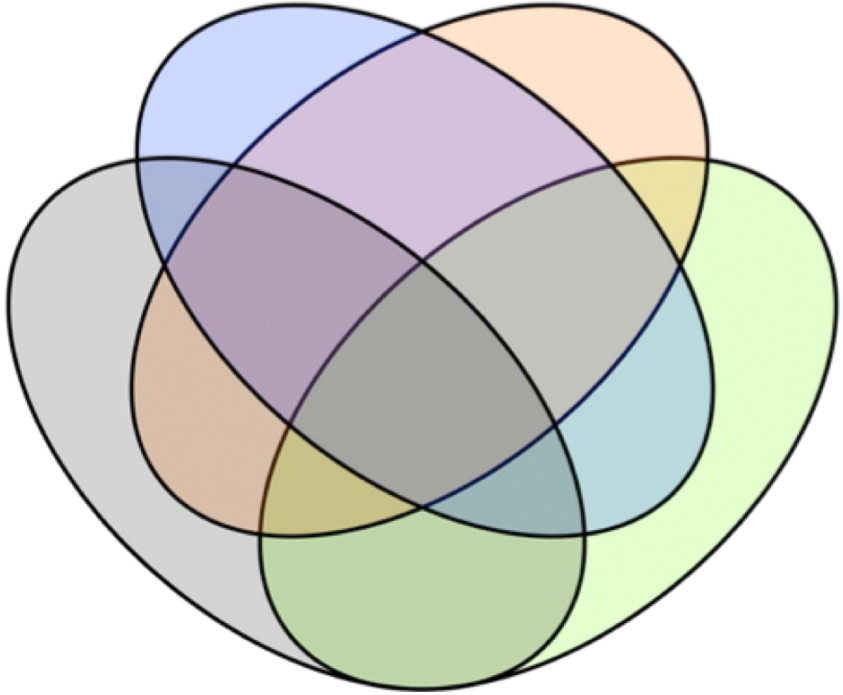
What are set-typed data?

Data items are often grouped into sets based on specific properties.

Venn diagrams

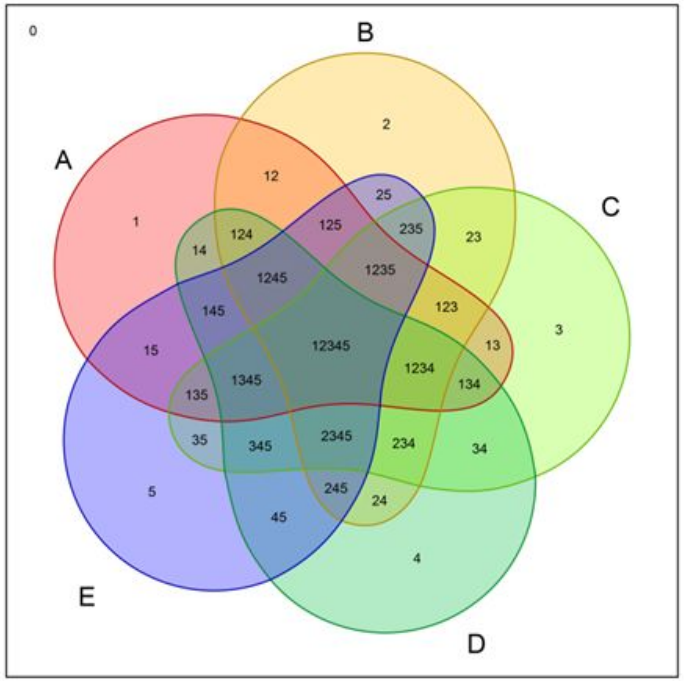
Show all possible set relations.

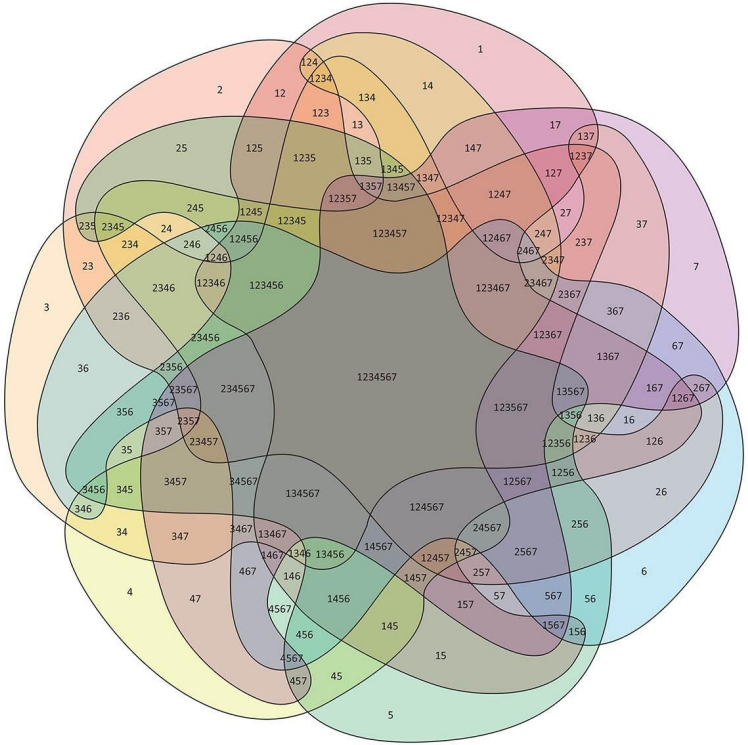




Venn diagrams

Get messy fast.



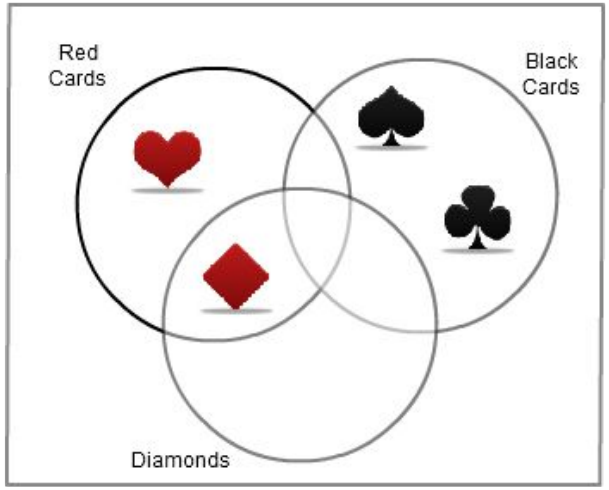


https://creately.com/blog/diagrams/venn-diagrams-vs-euler-diagrams/

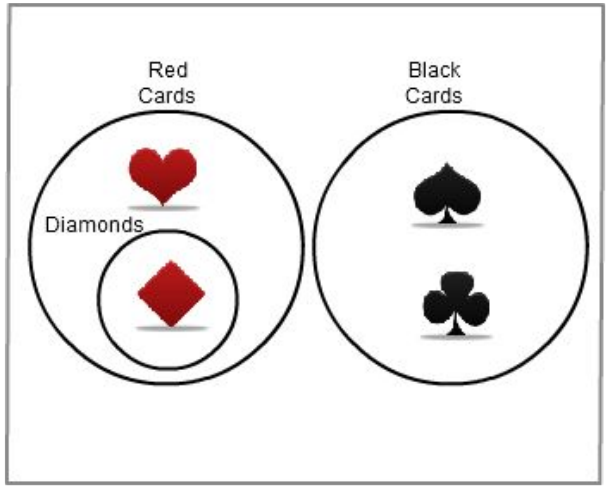
Euler diagrams

Only show existing set relations.

VENN



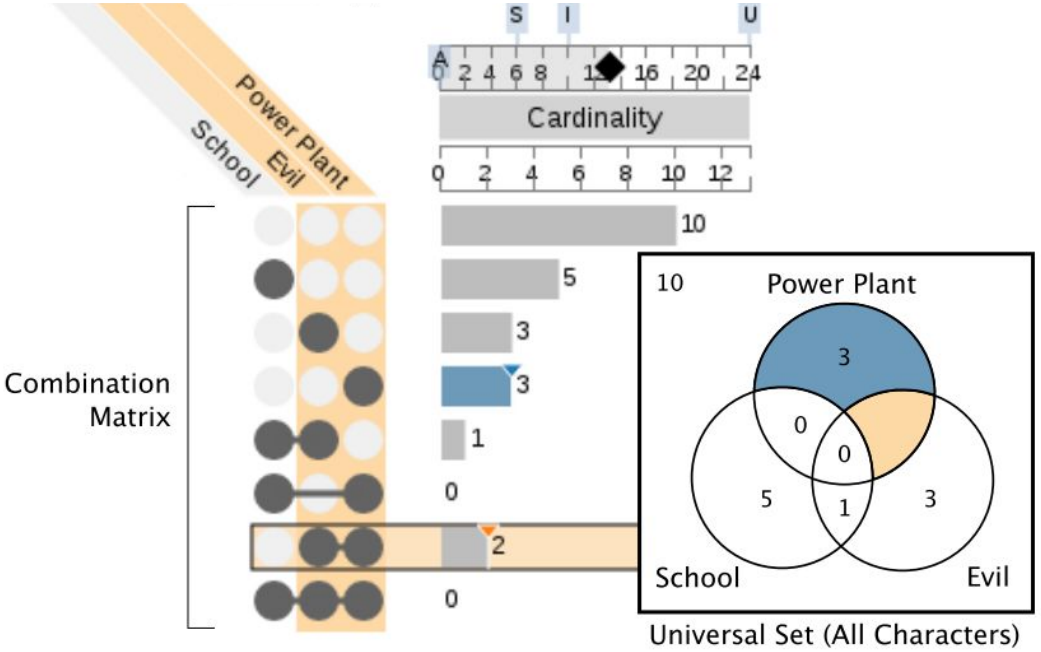
EULER



Matrix-based method -- UpSet

Row: set relation

Column: set



Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H., 2014. UpSet: visualization of intersecting sets. *IEEE transactions on visualization and computer graphics*, *20*(12), pp.1983-1992.

Recommended reading

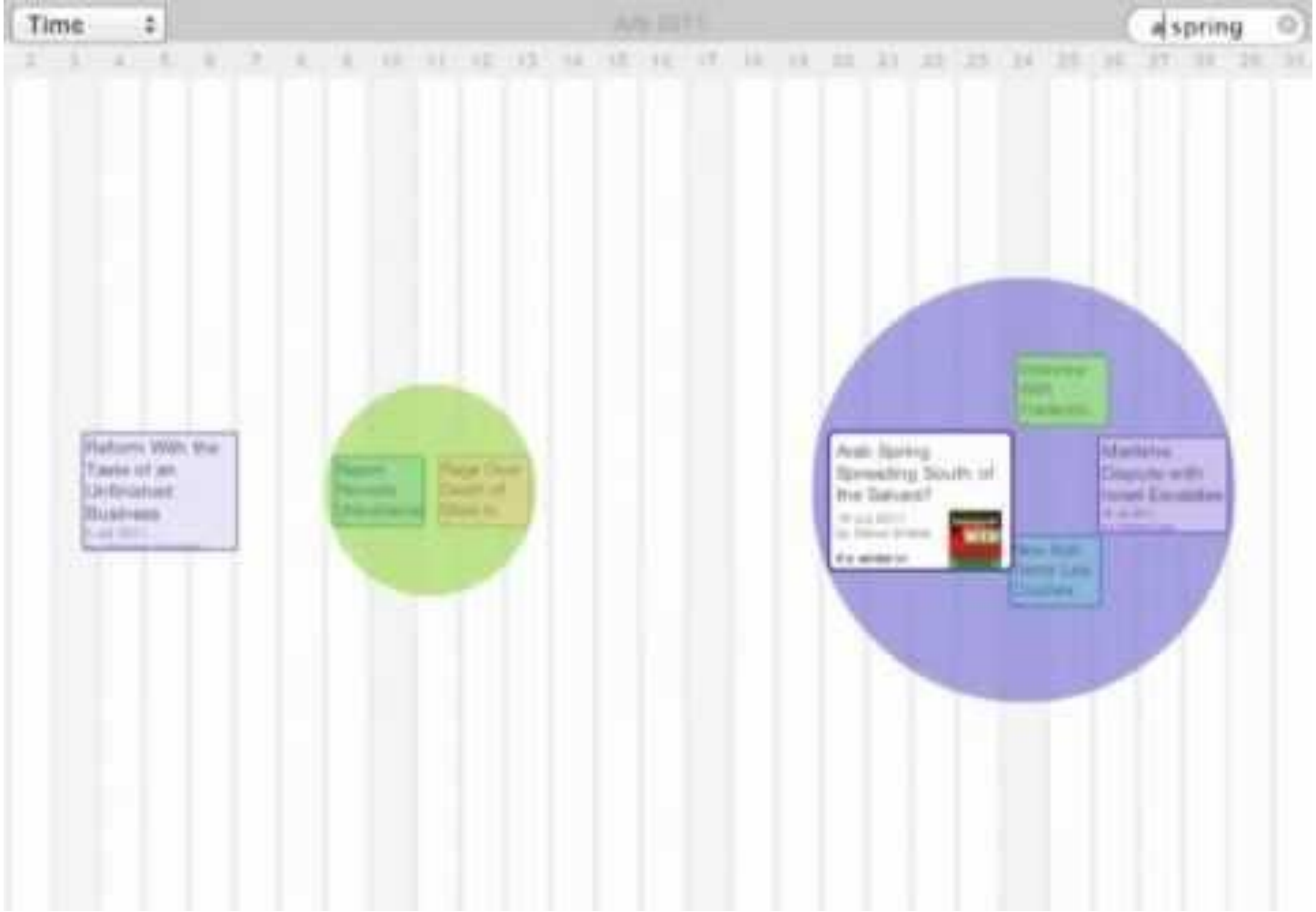
Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. Commun. Acm, 53(6), 59-67.
https://queue.acm.org/detail.cfm?searchterm=Mind+Maps&id=1805128

Interacting with visualizations

Shneiderman, B., 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization* (pp. 364-371).

Information seeking mantra

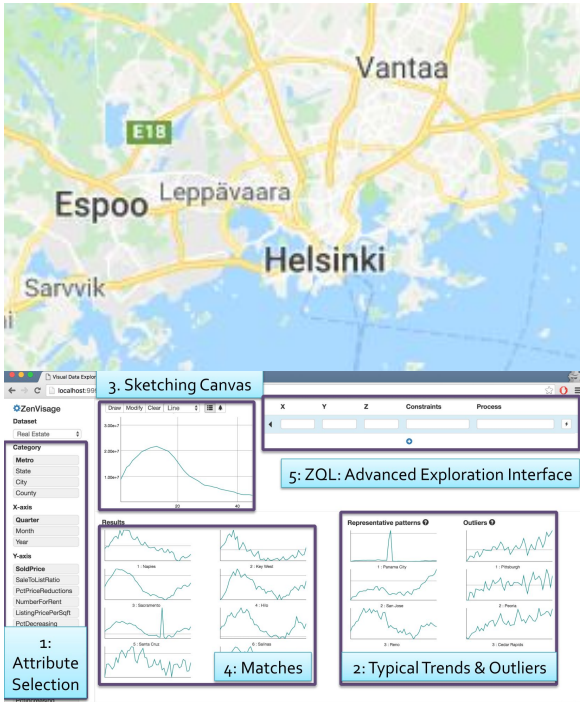
Overview first, zoom and filter, details on demand
Overview first, zoom and filter, details on demand
Overview first, zoom and filter, details on demand
Overview first, zoom and filter, details on demand
Overview first, zoom and filter, details on demand
Overview first, zoom and filter, details on demand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



Dörk, M., Carpendale, S., & Williamson, C. (2012, May). Fluid Views: a zoomable search environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 233-240). ACM.

Overview first, zoom and filter, details on demand
Though there can be exceptions…

E.g. when the system knows your context.
E.g. when the system jumps right into insights.



Recap

What is data visualization?

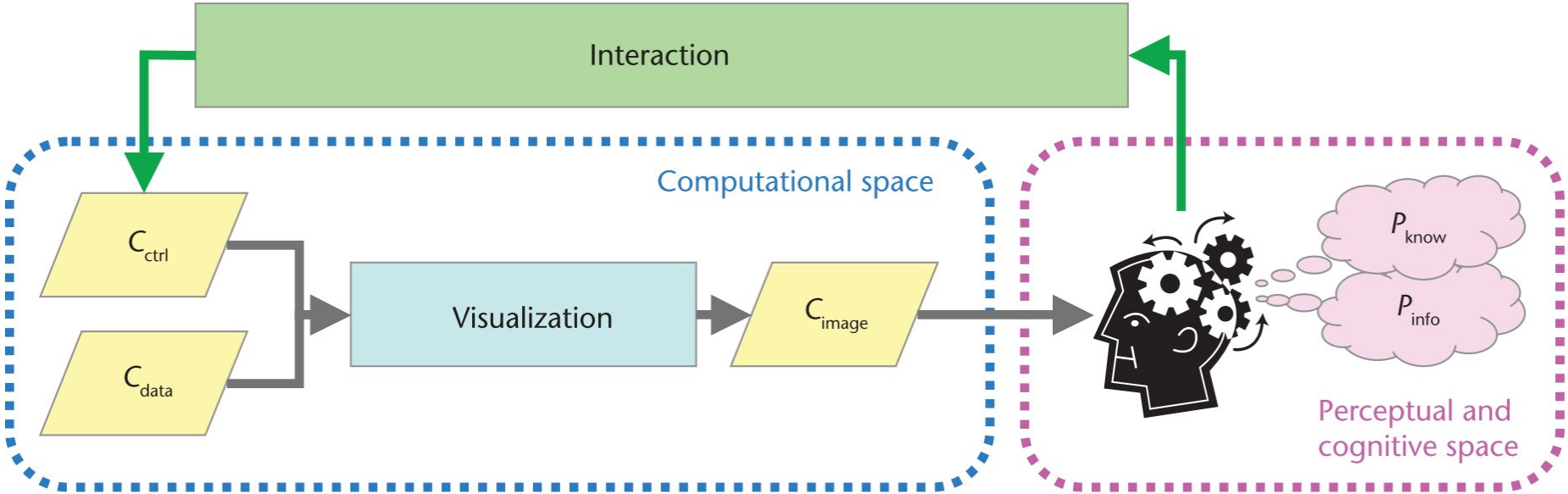
The use of computer-supported, interactive, visual representations of data to amplify cognition.

Purposes of visualization

To help

- Make a decision
- Tell a story
- Reasoning
- Discover knowledge
- ……

Recap -- Visualization process



Recap -- Examples visualizing various types of data

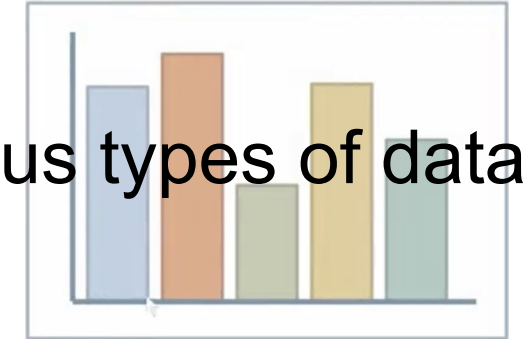
Two-dimensional data

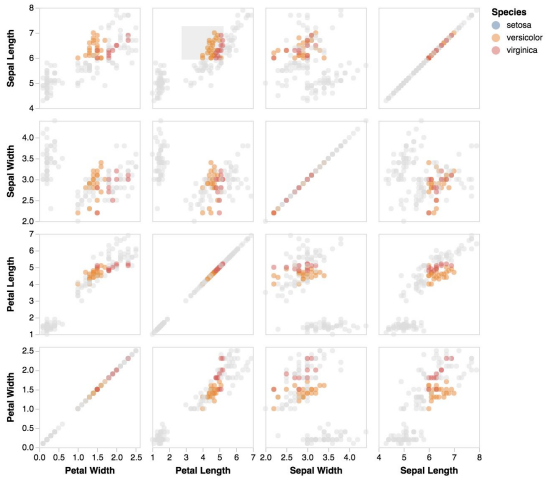
Multi-dimensional data

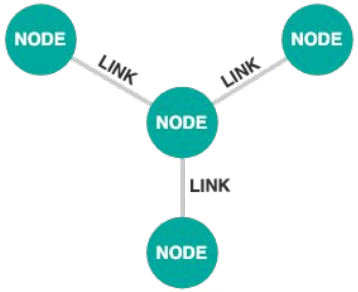
Graph

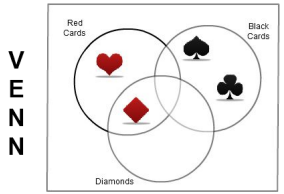
Hierarchical data

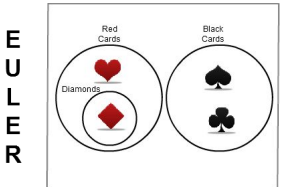
Set-typed data

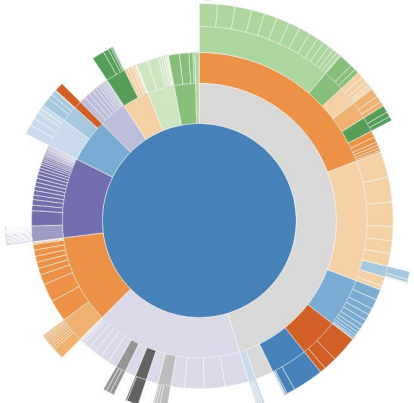








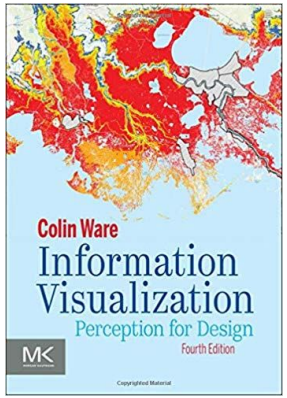




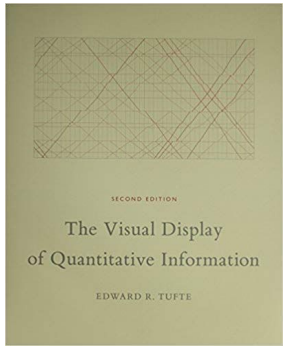
Recap -- Information seeking mantra

Overview first, zoom and filter, details on demand

Course textbooks

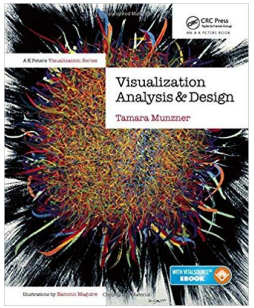


Colin Ware. **Information Visualization: Perception for Design (Interactive Technologies)**, 4th Edition. Morgan Kaufmann, 2020. [ebook available]

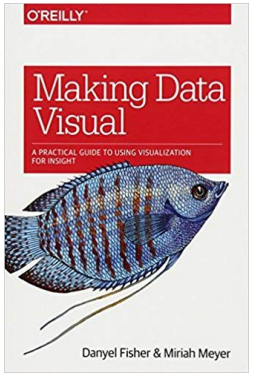


Edward Tufte. **The Visual Display of Quantitative Information**, 2nd Edition. Graphics Press, 2001.

Course textbooks



Tamara Munzner. **Visualization Analysis and Design**. CRC Press, 2014. [ebook available]



Miriah Meyer, Danyel Fisher. **Making Data Visual: A Practical Guide to Using Visualization for Insight**. O'Reilly Media, 2018.

Tools to create visualization

Tableau https://www.tableau.com/

No limit on the number of data points; interactions have delay.

Power BI https://powerbi.microsoft.com/en-au/

Thousands of rows depend on chart types.

Plotly https://plot.ly/

25,000 x-y points.

Vega-Lite https://vega.github.io/vega-lite/

Limit to 10,000 data points.

D3.js https://d3js.org/

1000 or so data points for smooth interaction; use canvas to increase to 10,000.