

Data visualization

An introduction

Chen He
chen.he@helsinki.fi

Why do we need visualization?

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

mean of the x values = 9.0

mean of the y values = 7.5

equation of the least-squared regression line is: $y = 3 + 0.5x$

sums of squared errors (about the mean) = 110.0

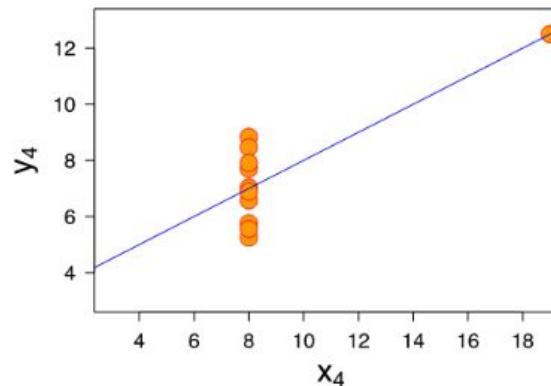
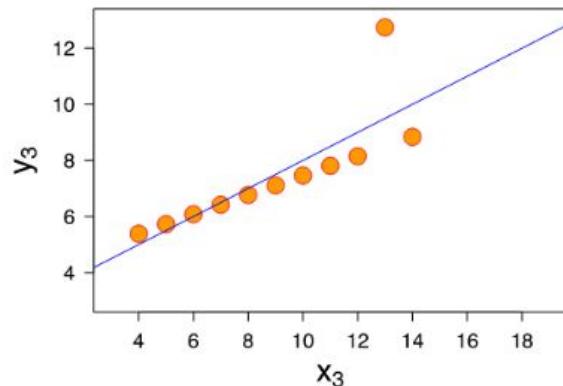
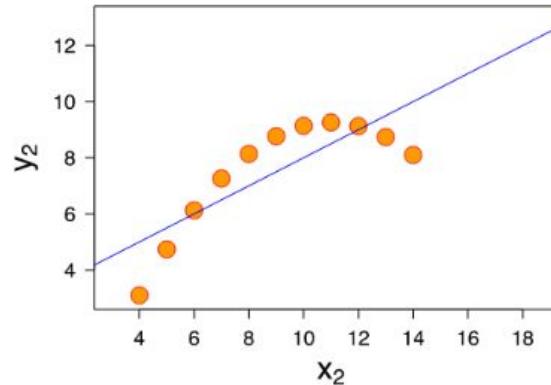
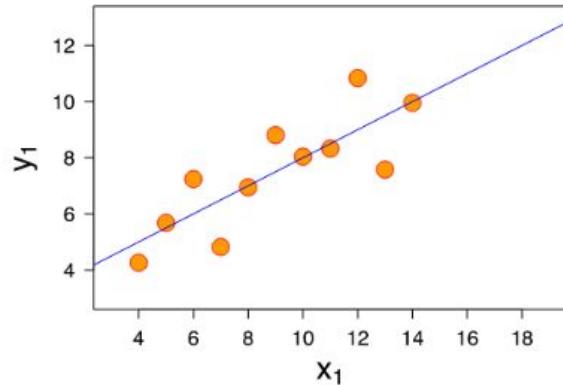
regression sums of squared errors (variance accounted for by x) = 27.5

residual sums of squared errors (about the regression line) = 13.75

correlation coefficient = 0.82

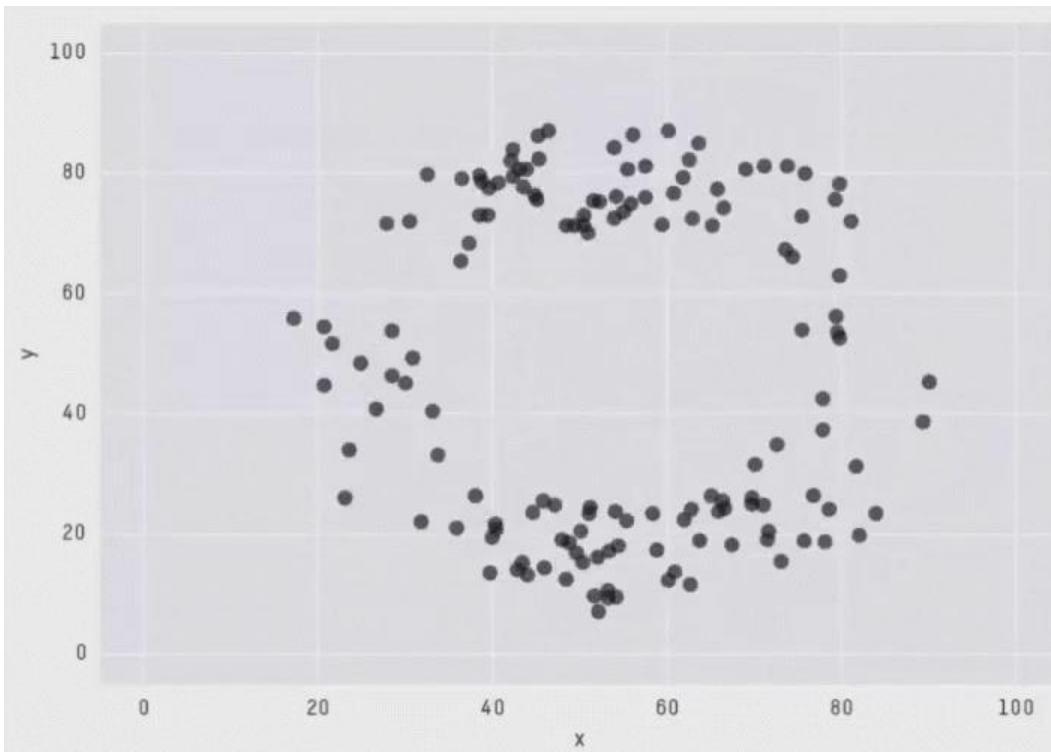
coefficient of determination = 0.67

Why do we need visualization?



Anscombe's Quartet

Why do we need visualization?



X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06

What is **visualization**?

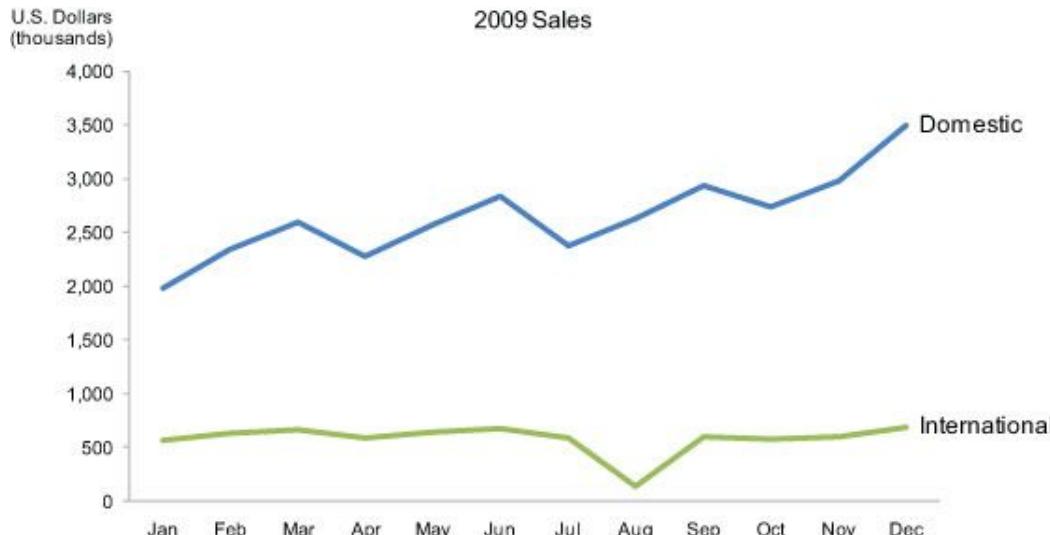
The use of computer-supported, interactive, visual representations of data to amplify cognition.

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

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

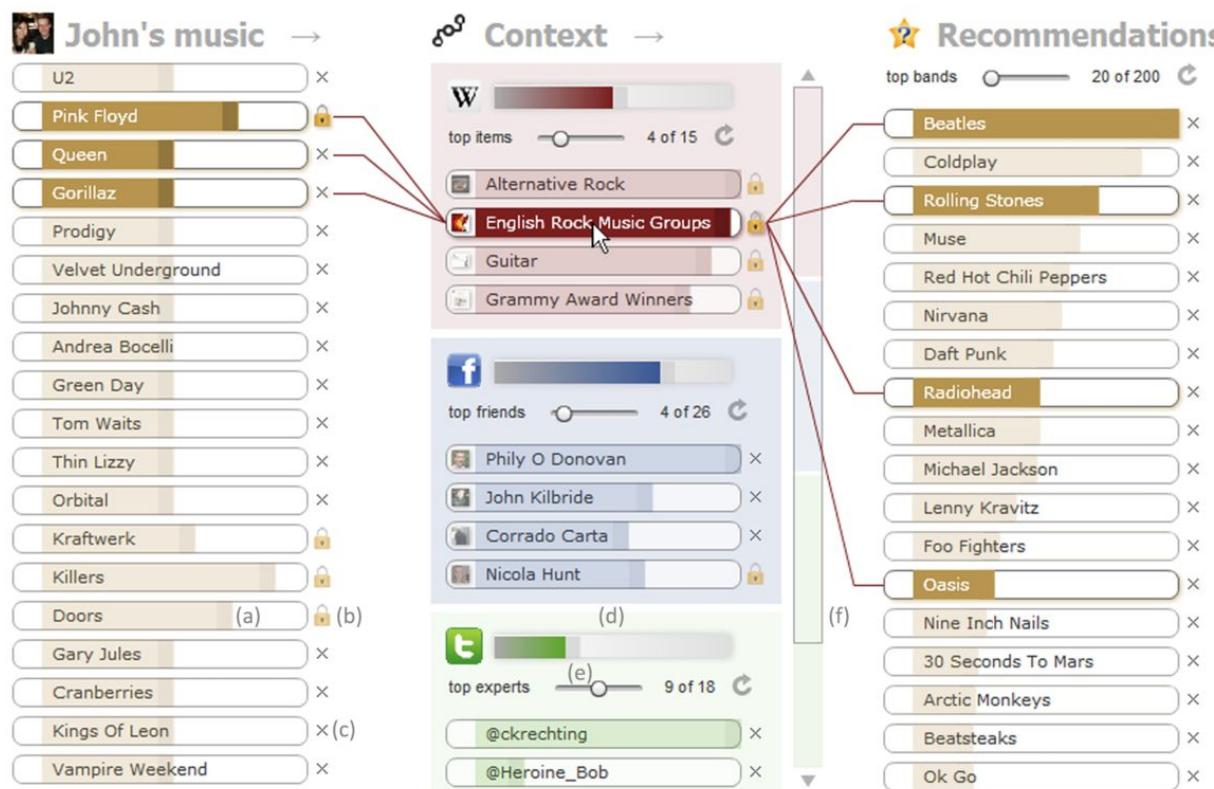


Purposes of visualization

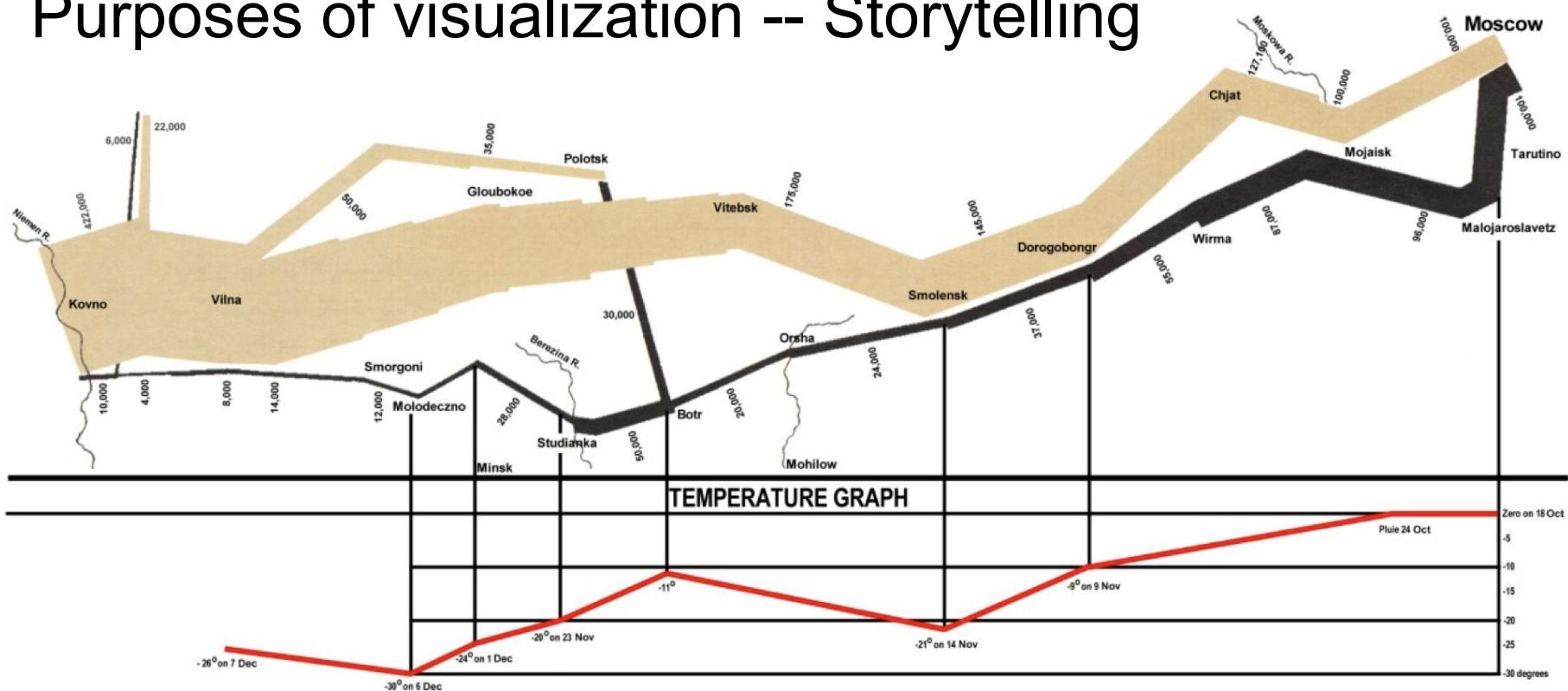
To help

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

Purposes of visualization -- Decision making

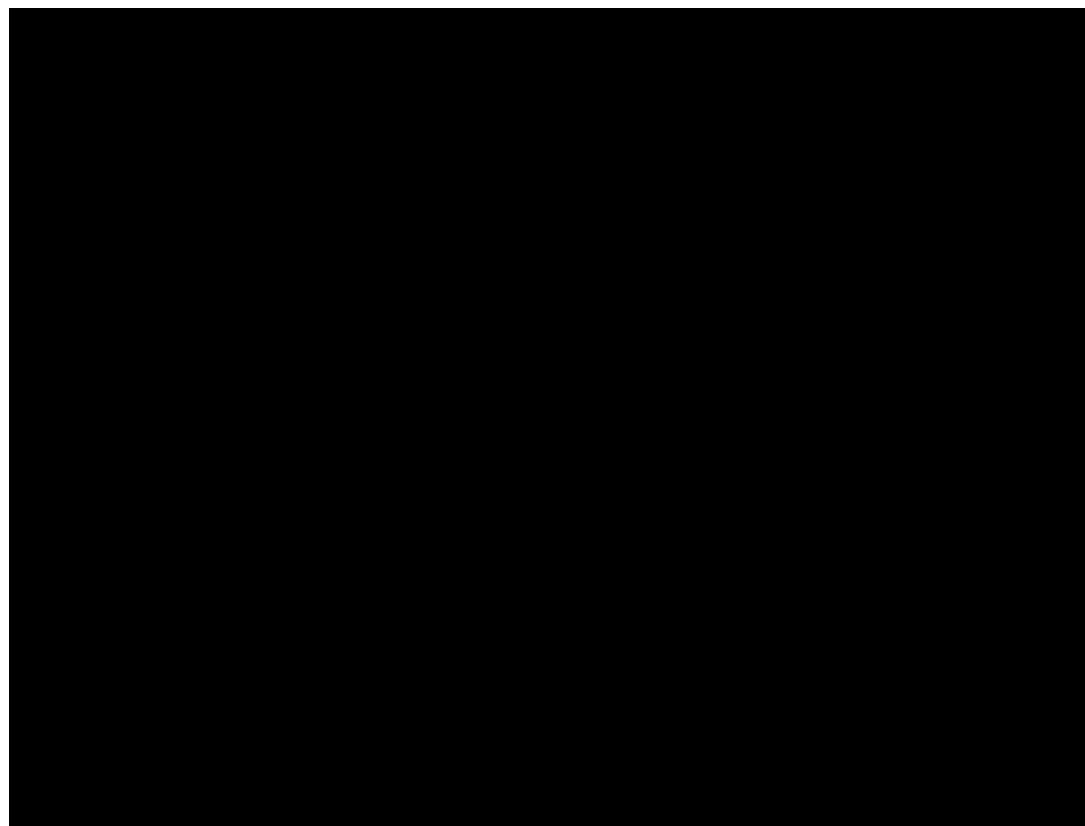


Purposes of visualization -- Storytelling

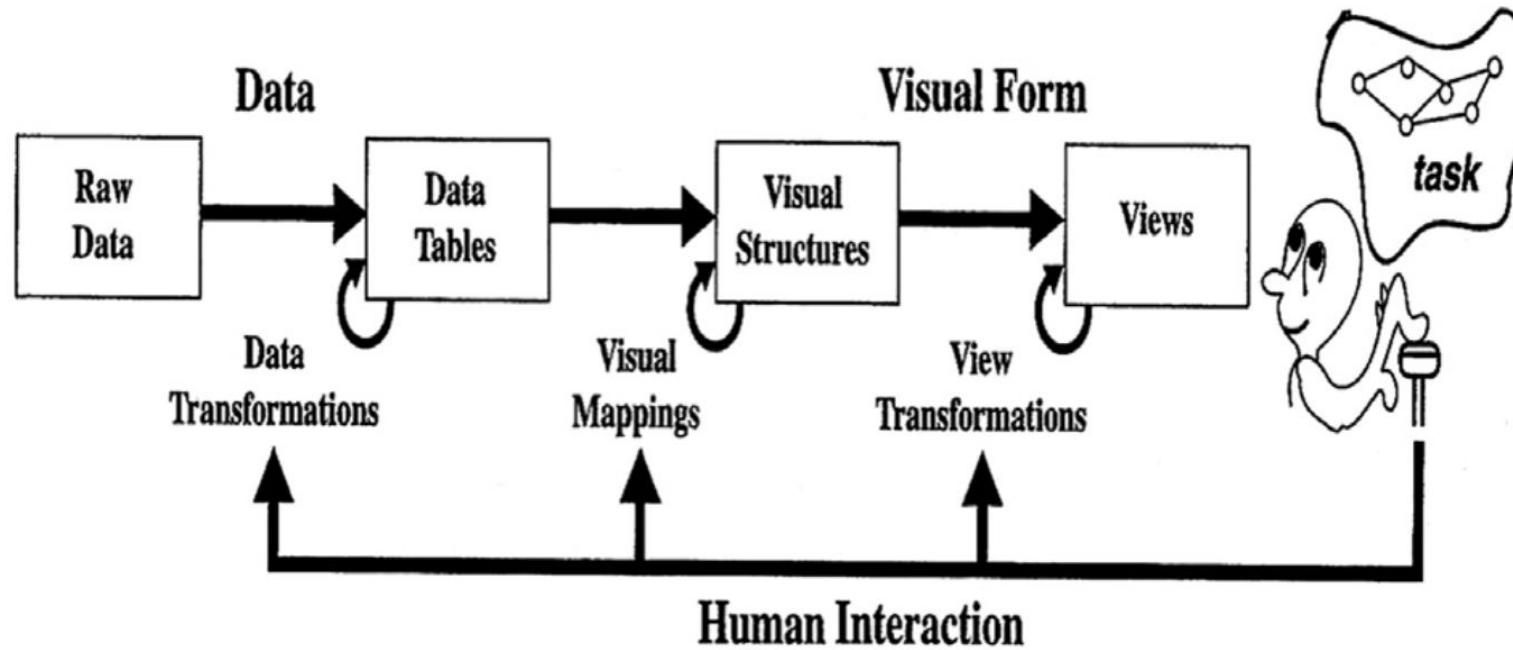


Dork, M., Gruen, D., Williamson, C. and Carpendale, S., 2010. A visual backchannel for large-scale events. IEEE transactions on visualization and computer graphics, 16(6), pp.1129-1138.

Purposes of visualization -- Exploration & discovery



Visualization framework



Visualizing different types of data

Two-dimensional data

Multi-dimensional data

Graph

Hierarchical data

Set-typed data

Visualizing different types of data

Two-dimensional data

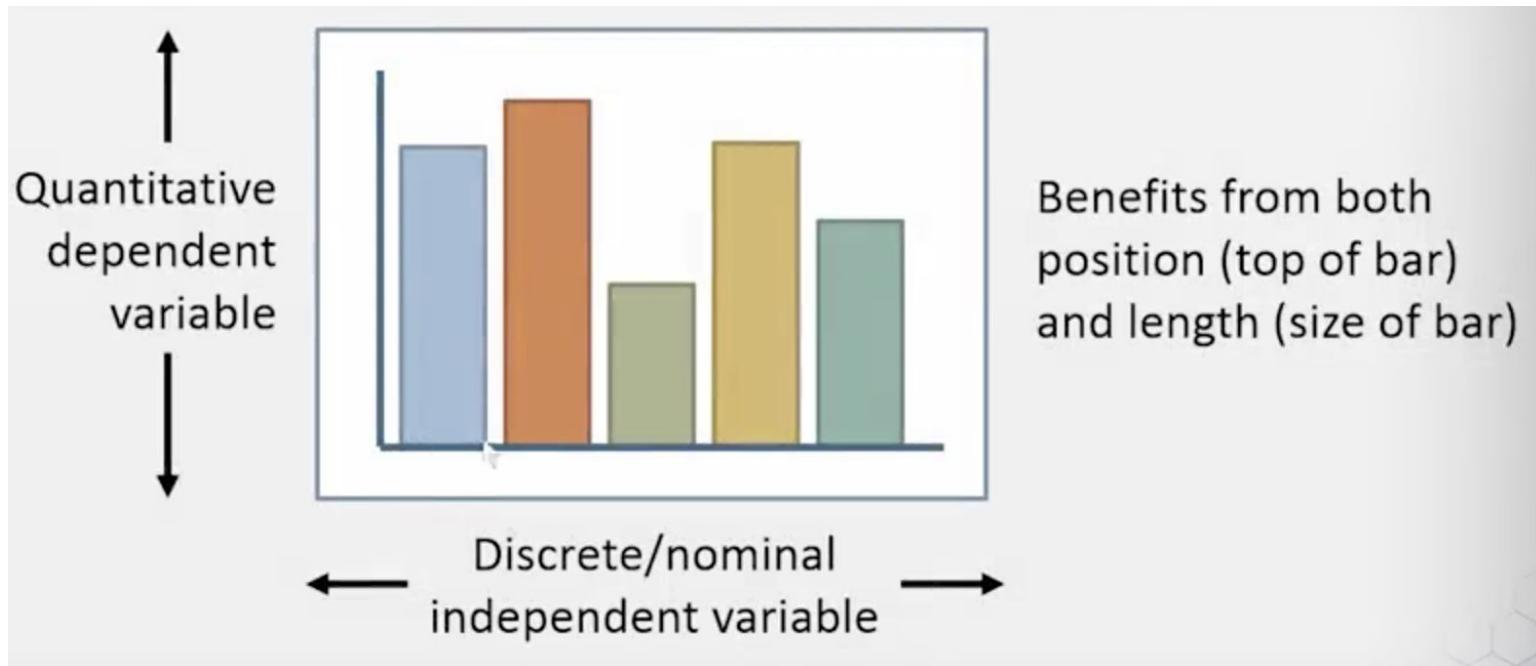
Multi-dimensional data

Graph

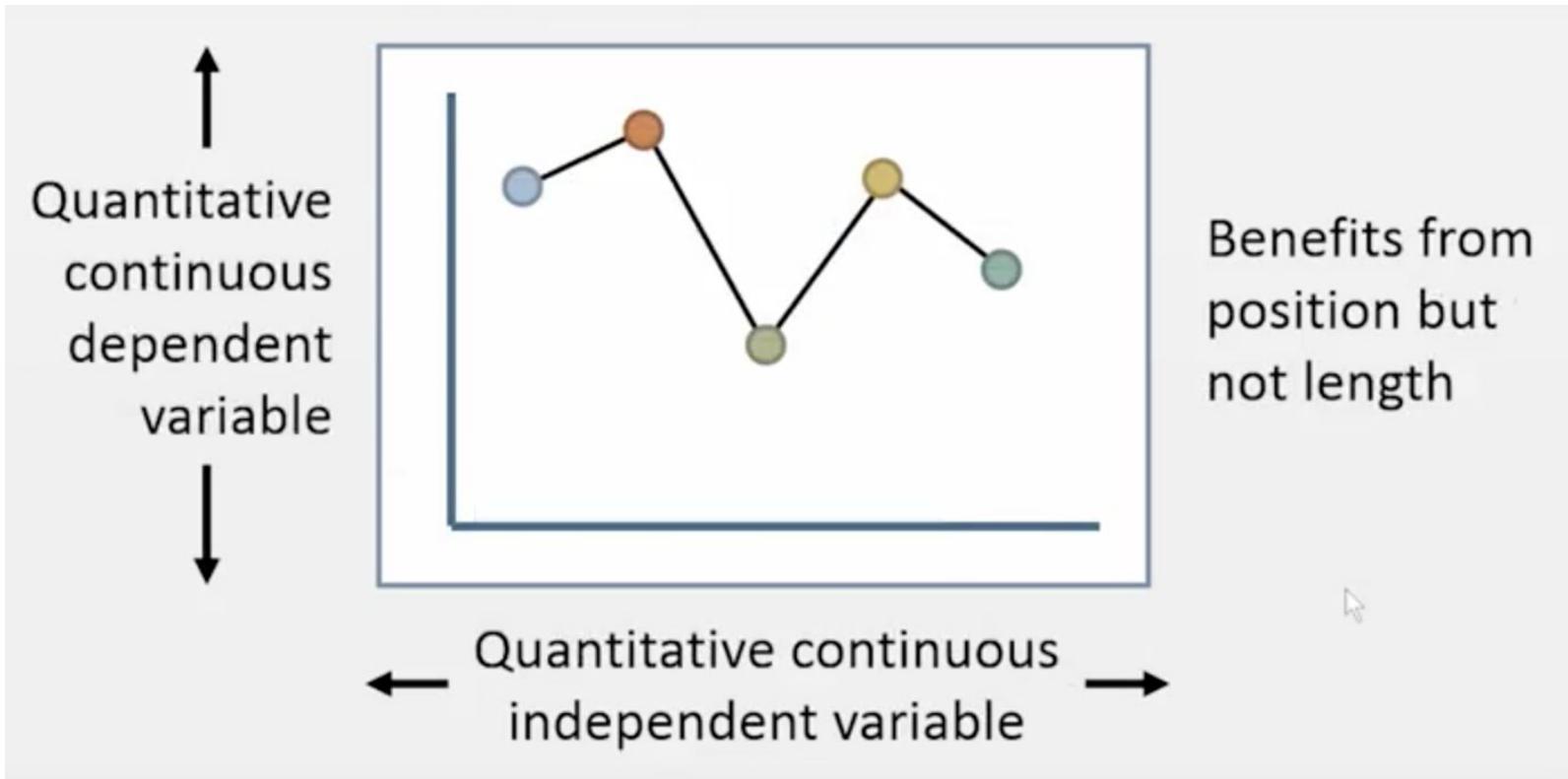
Hierarchical data

Set-typed data

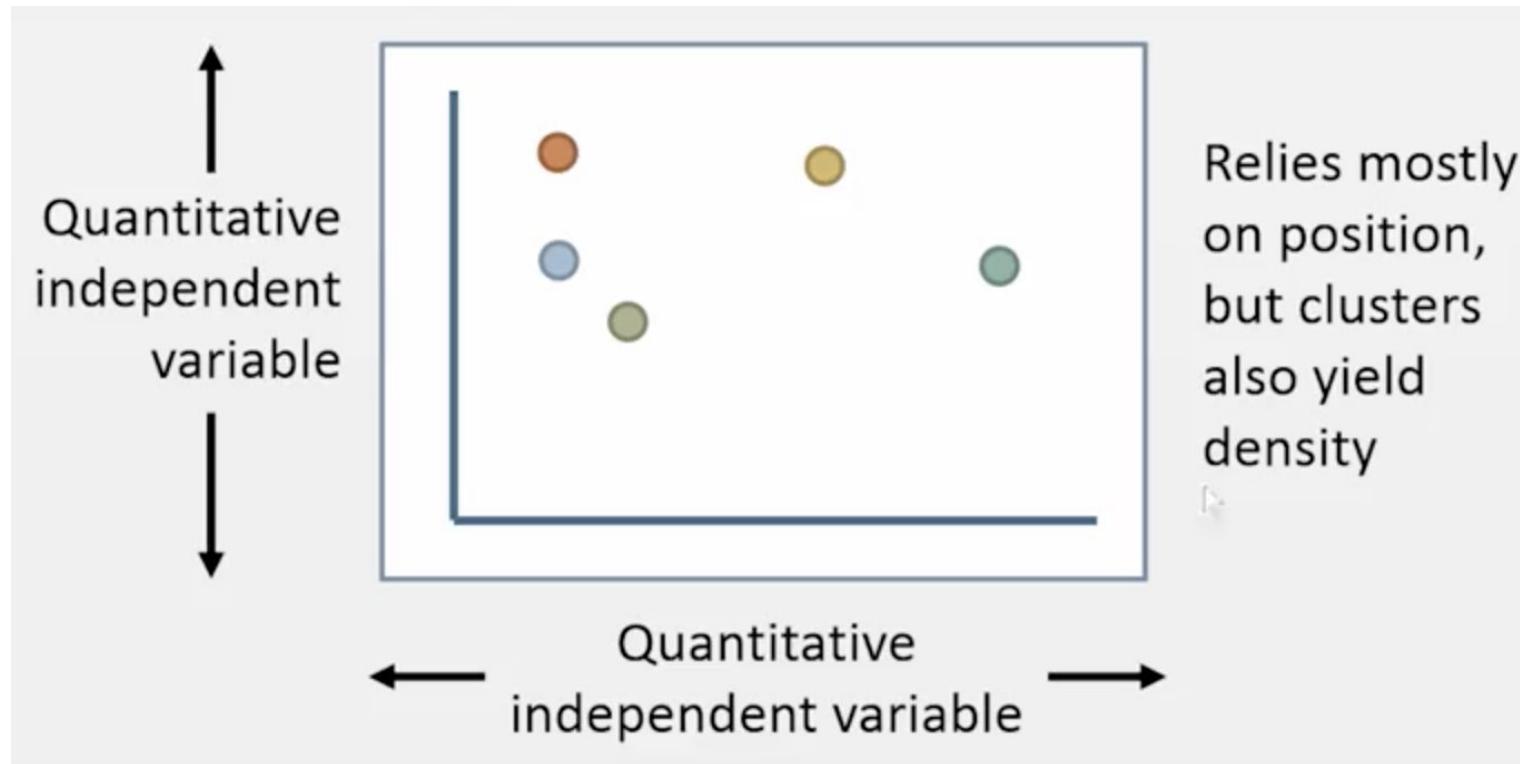
Two-dimensional data -- Bar chart



Two-dimensional data -- Line chart



Two-dimensional data -- Scatterplot



Visualizing different types of data

Two-dimensional data

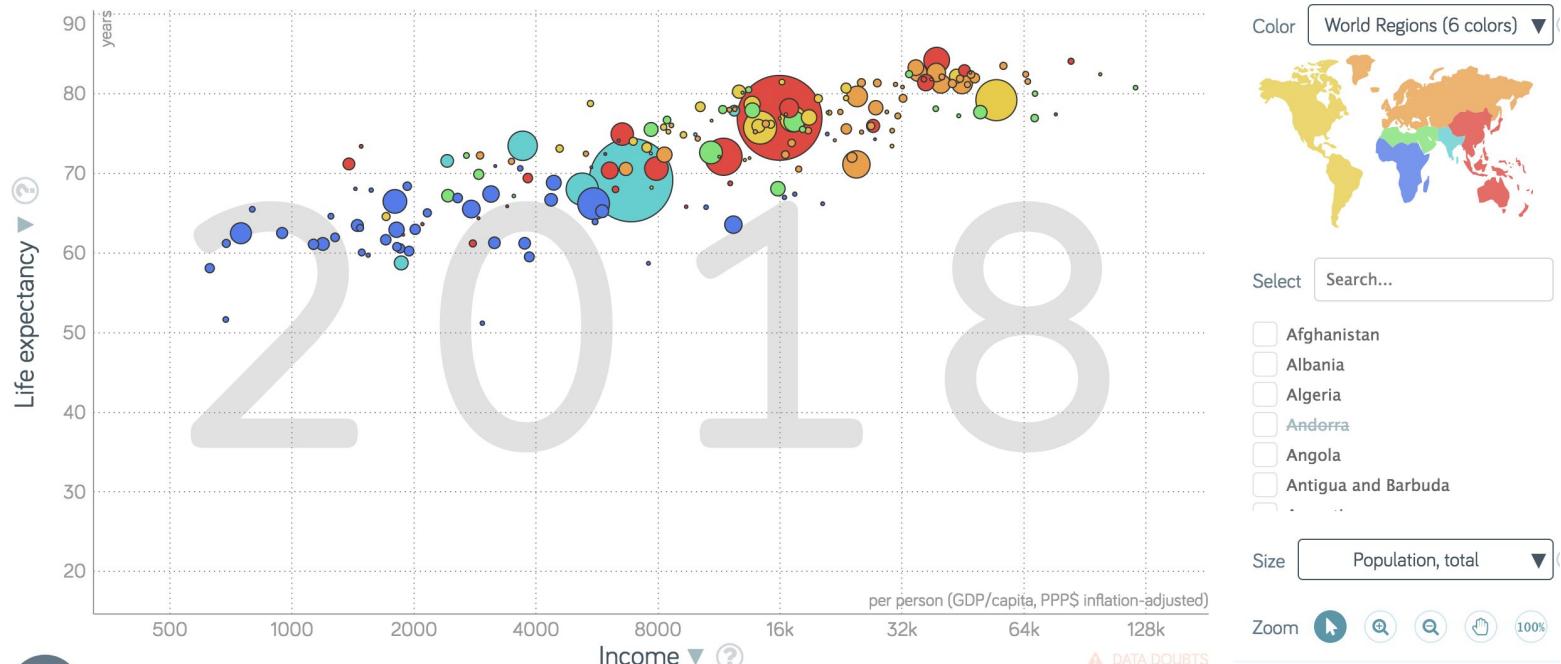
Multi-dimensional data

Graph

Hierarchical data

Set-typed data

Add additional dimensions on top of 2D charts



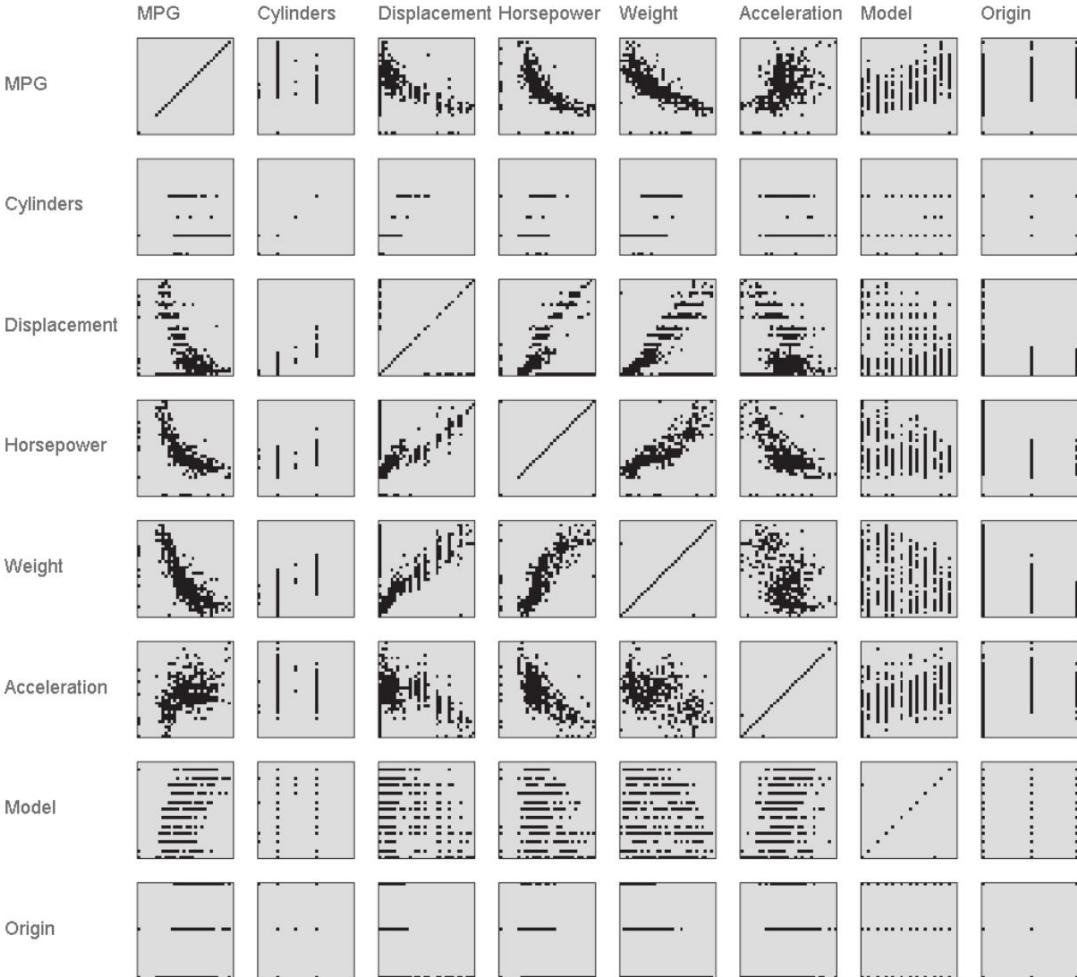
Scatterplot matrix

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.



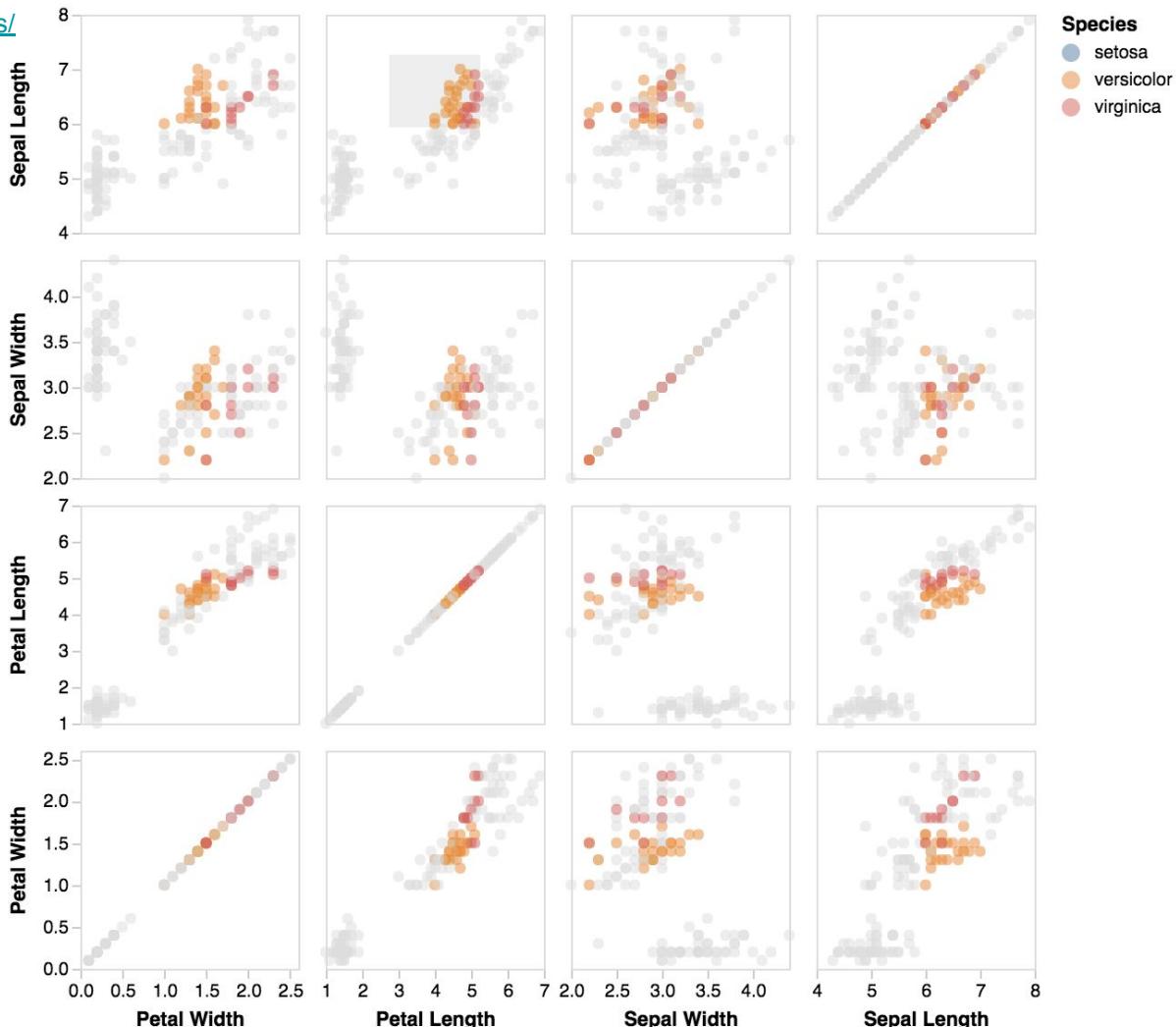
Interacting with multi-dimensional data

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.

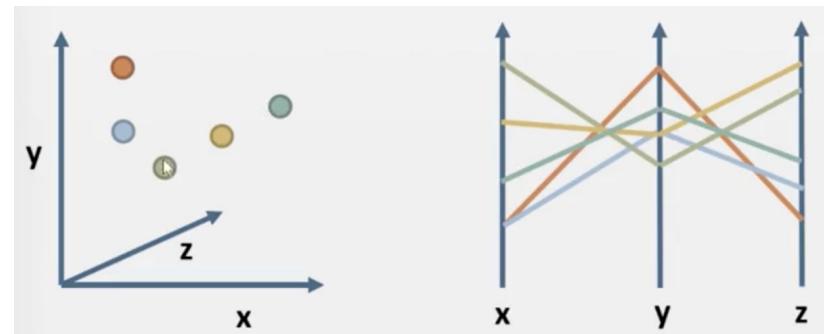
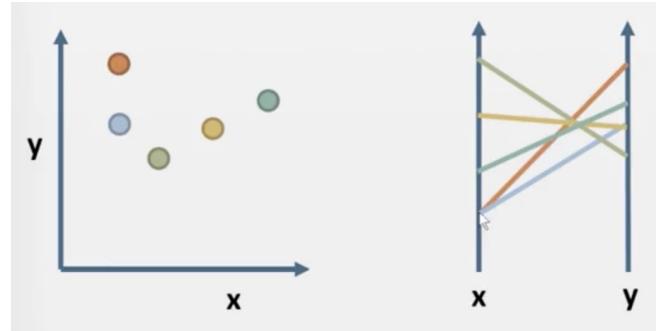
<https://vega.github.io/vega/examples/brushing-scatter-plots/>

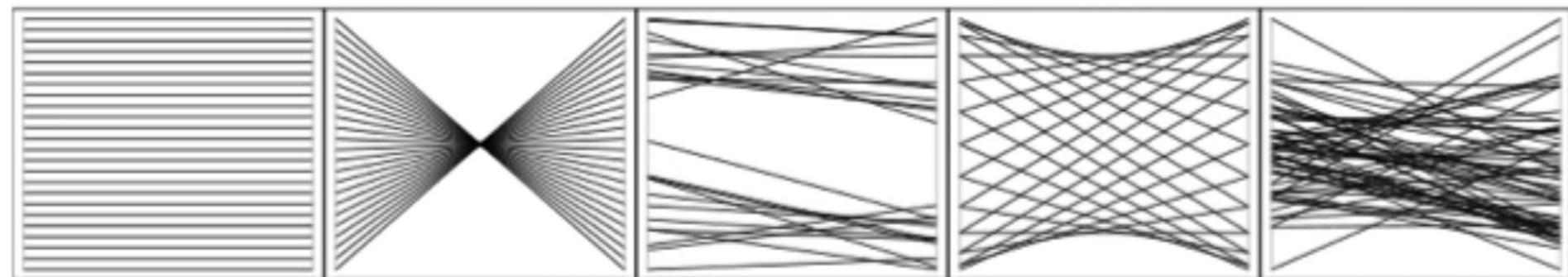
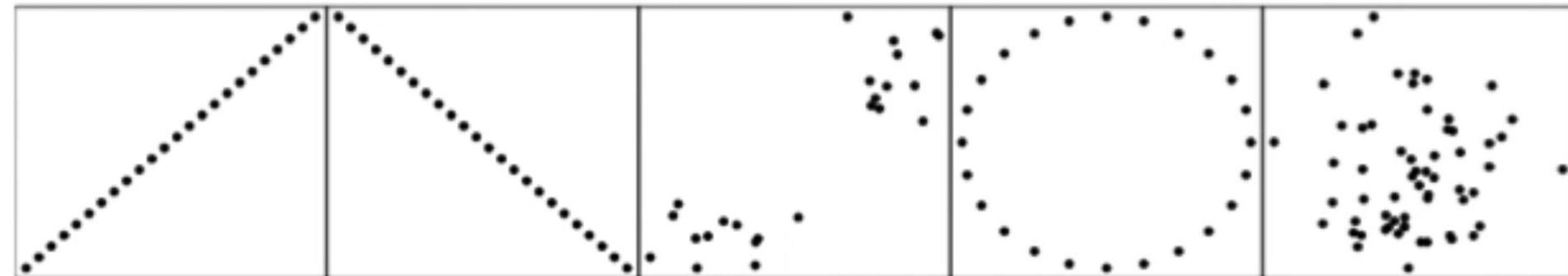
Brushing & linking scatter plots



Multi-dimensional data -- parallel coordinates

<https://www.coursera.org/learn/datavisualization/lecture/2v40S/2-2-2-parallel-coordinates>

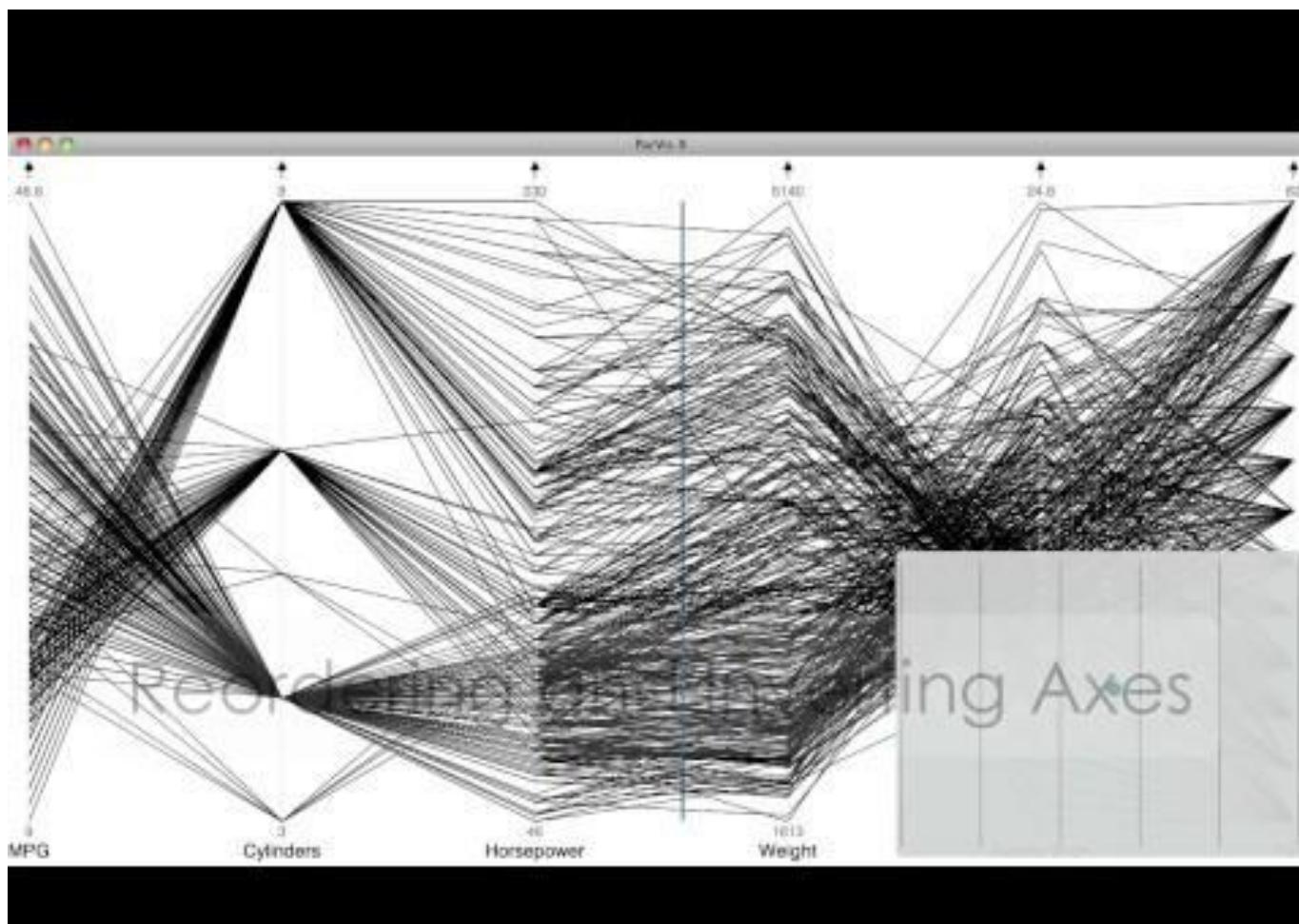




Interacting with parallel coordinates

<https://vimeo.com/13437693>

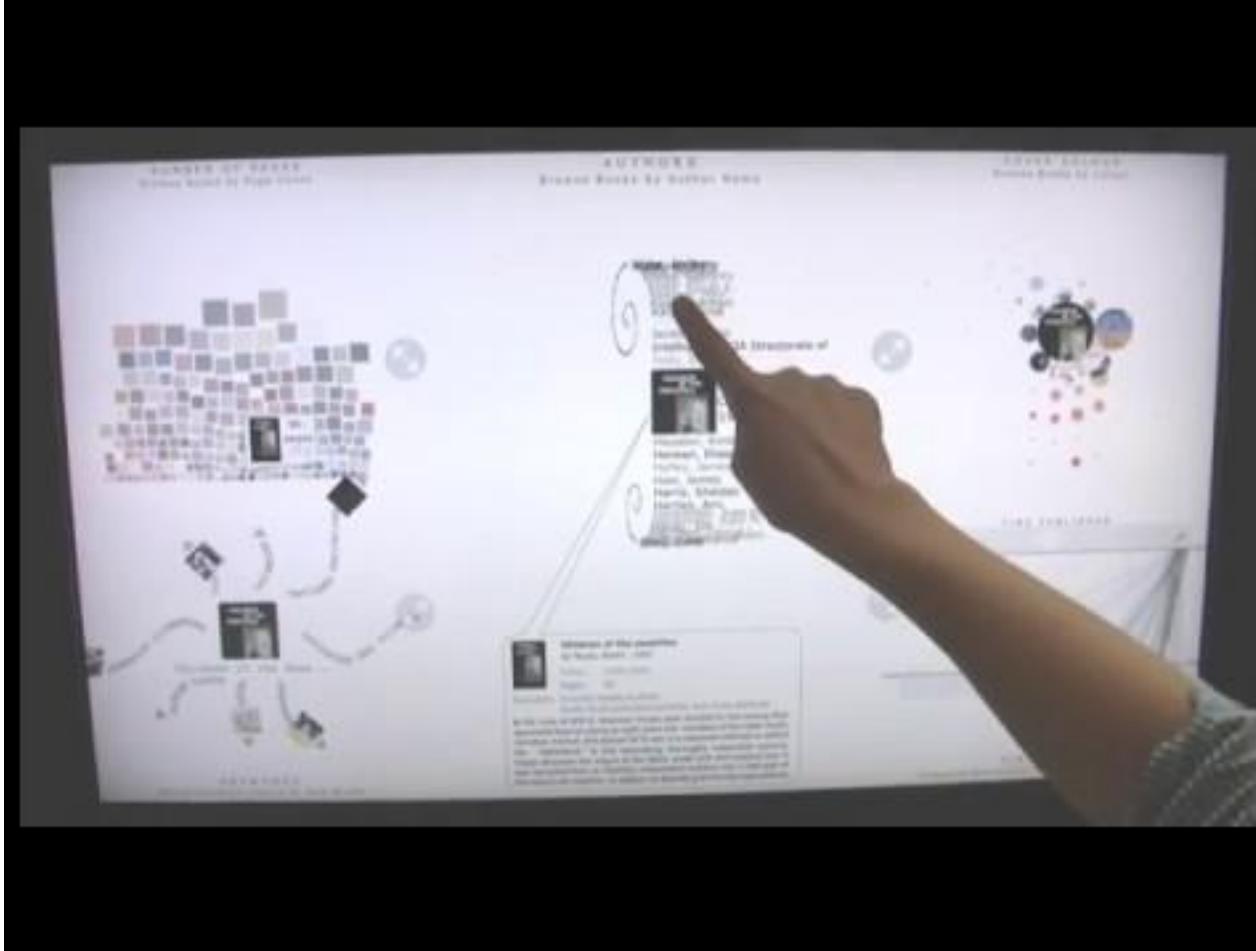
Sort an axis;
Reorder axes;
Single axis brushing;
Angular brushing.



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.

Visualizing different types of data

Two-dimensional data

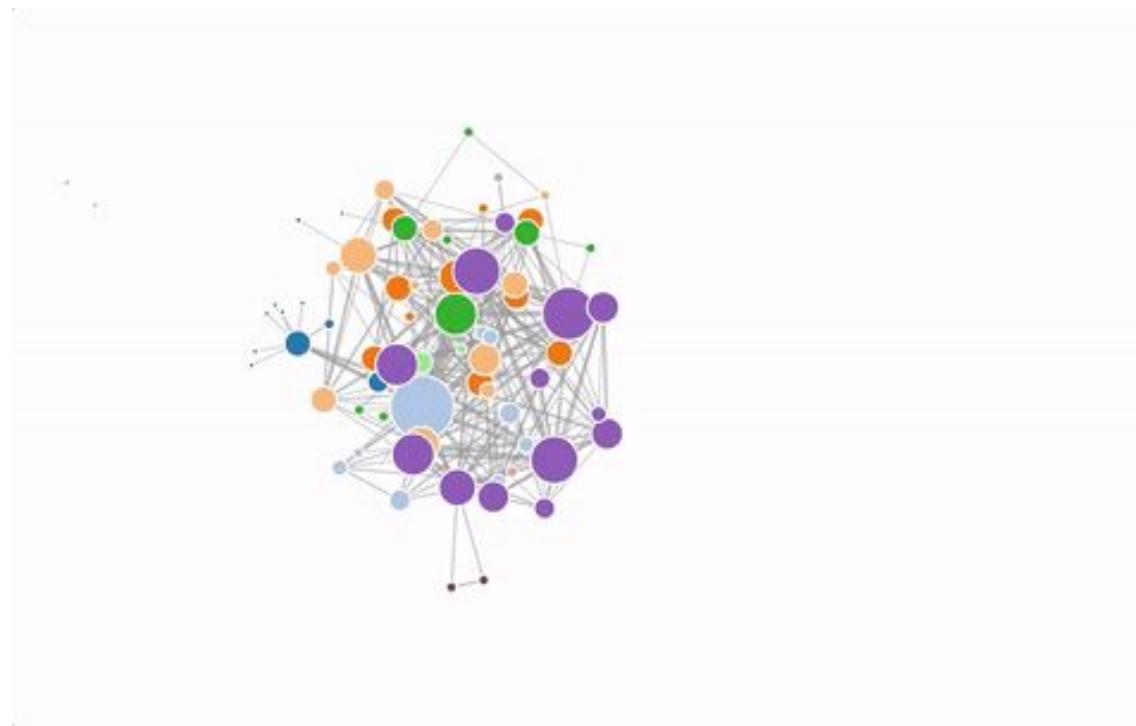
Multi-dimensional data

Graph

Hierarchical data

Set-typed data

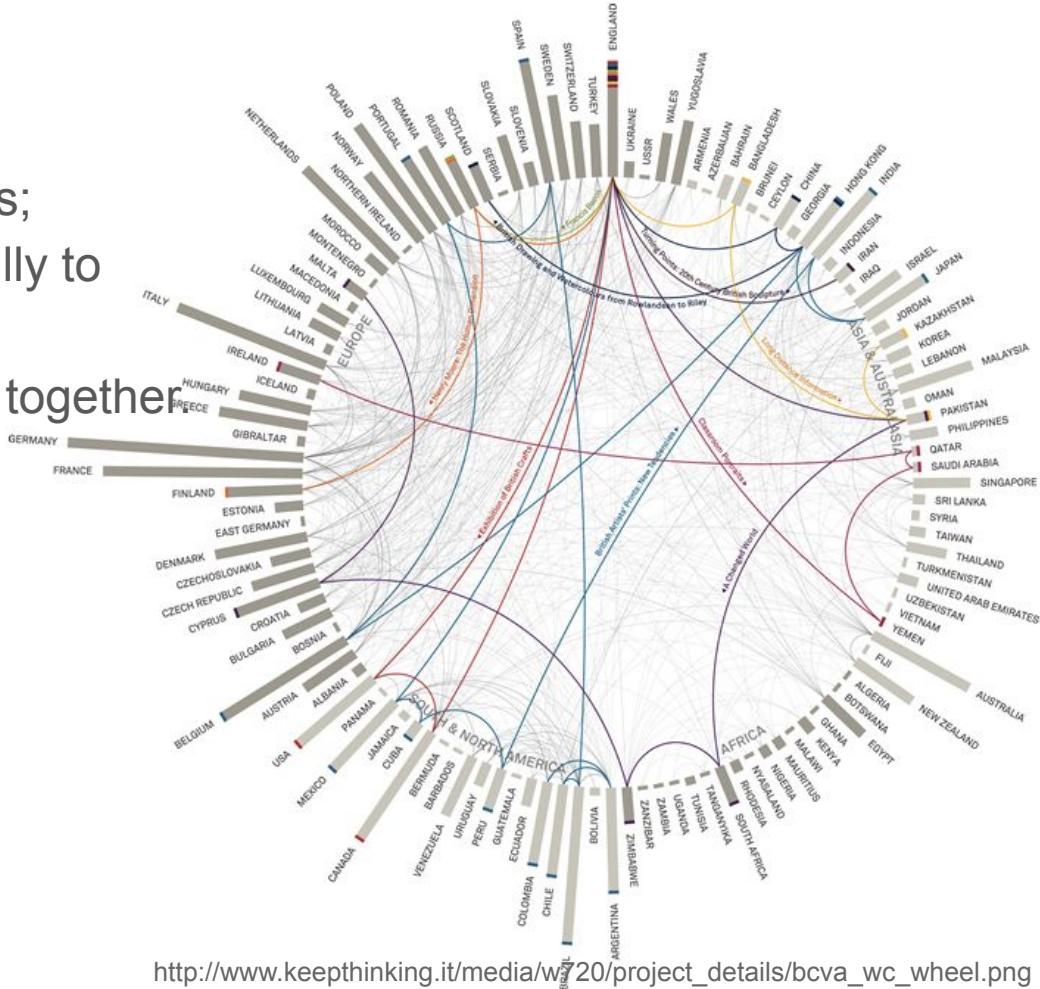
Graph -- Force-directed layout



<https://bl.ocks.org/heybignick/3faf257bbbb7743bb72310d03b86ee8>

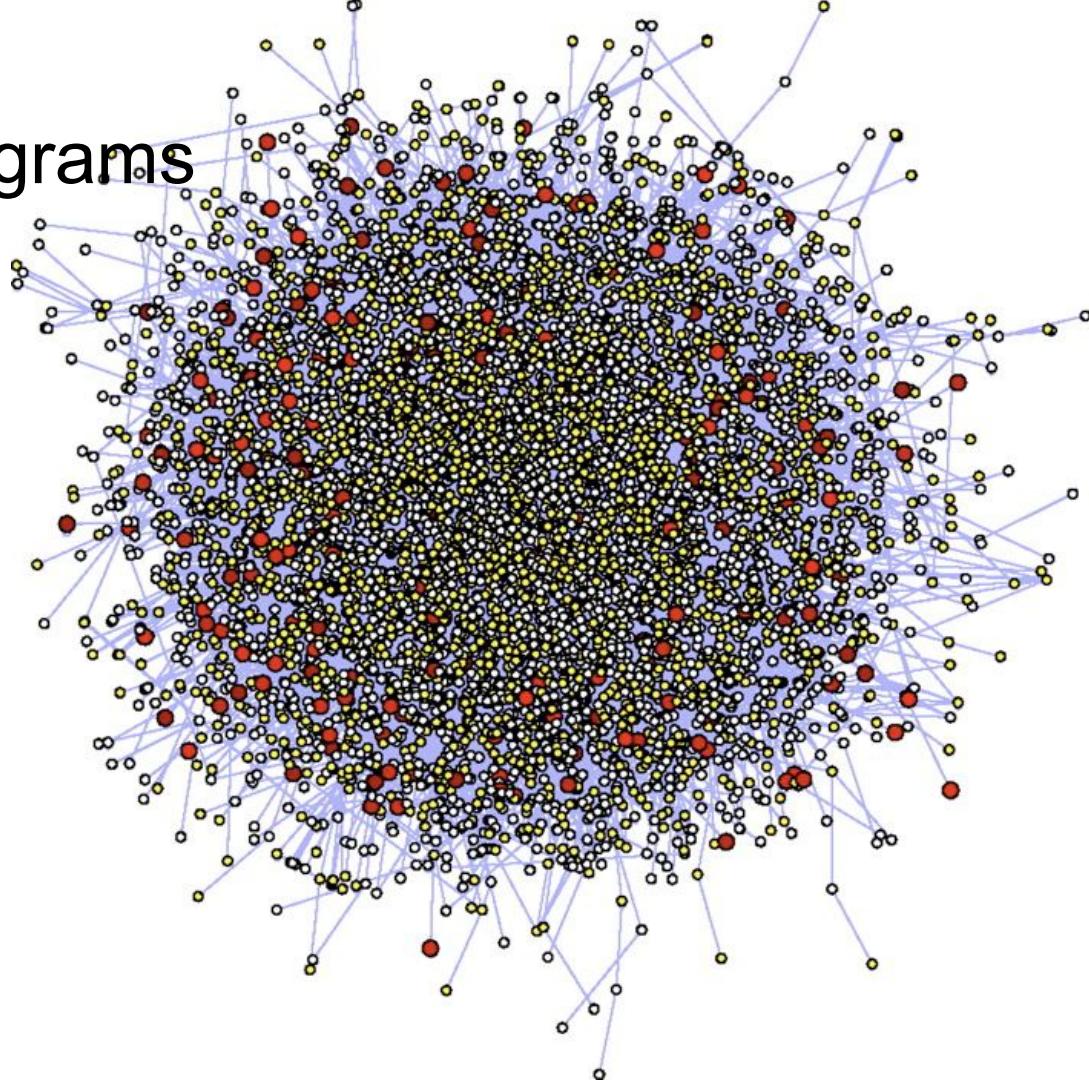
Graph -- Circular layout

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

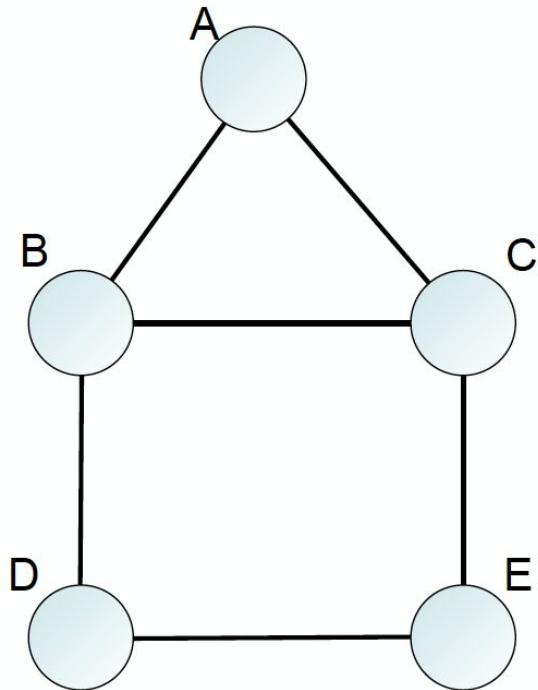


Graph -- Node-link diagrams

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

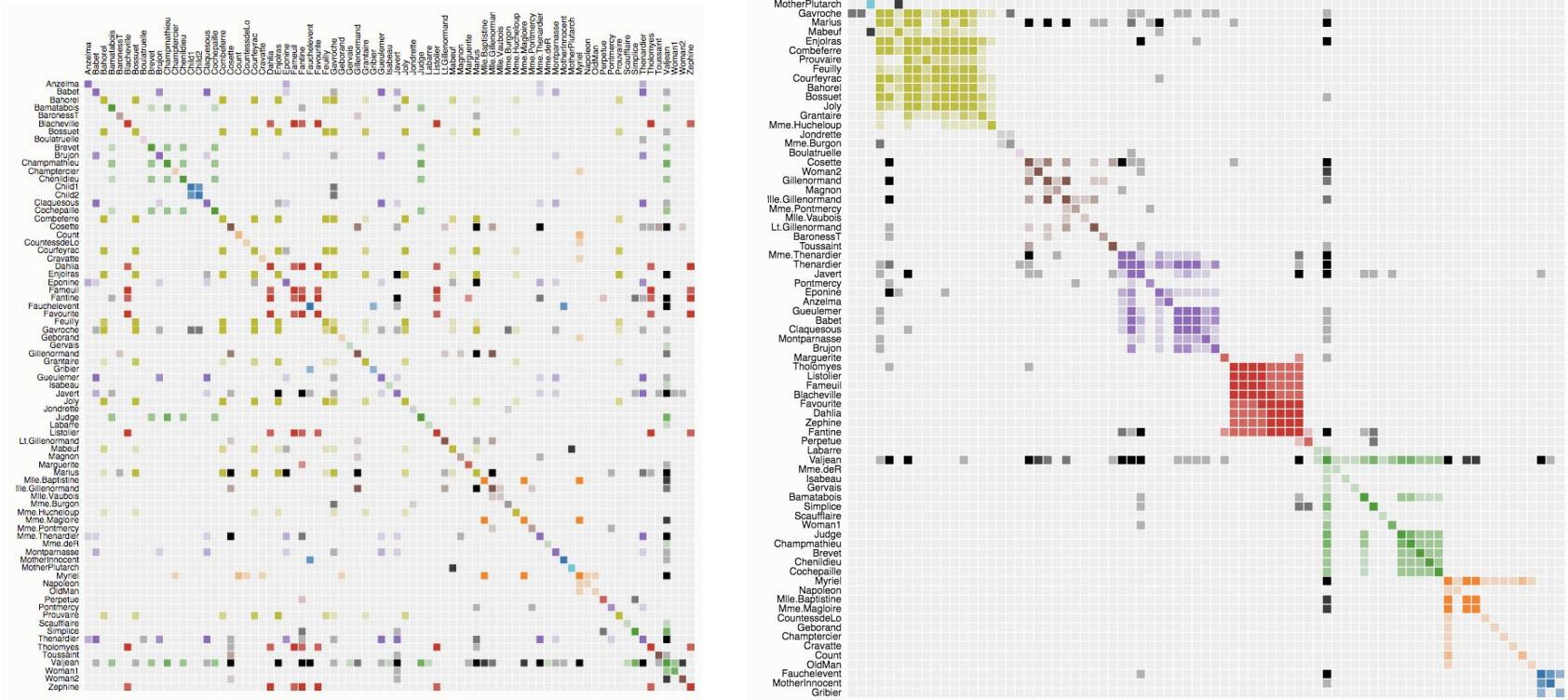


Graph -- Adjacency matrix



	A	B	C	D	E
A					
B					
C					
D					
E					

Graph -- Adjacency matrix



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

Graph -- Adjacency matrix

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

	A	B	C	D	E
A					
B					
C					
D					
E					

Visualizing different types of data

Two-dimensional data

Multi-dimensional data

Graph

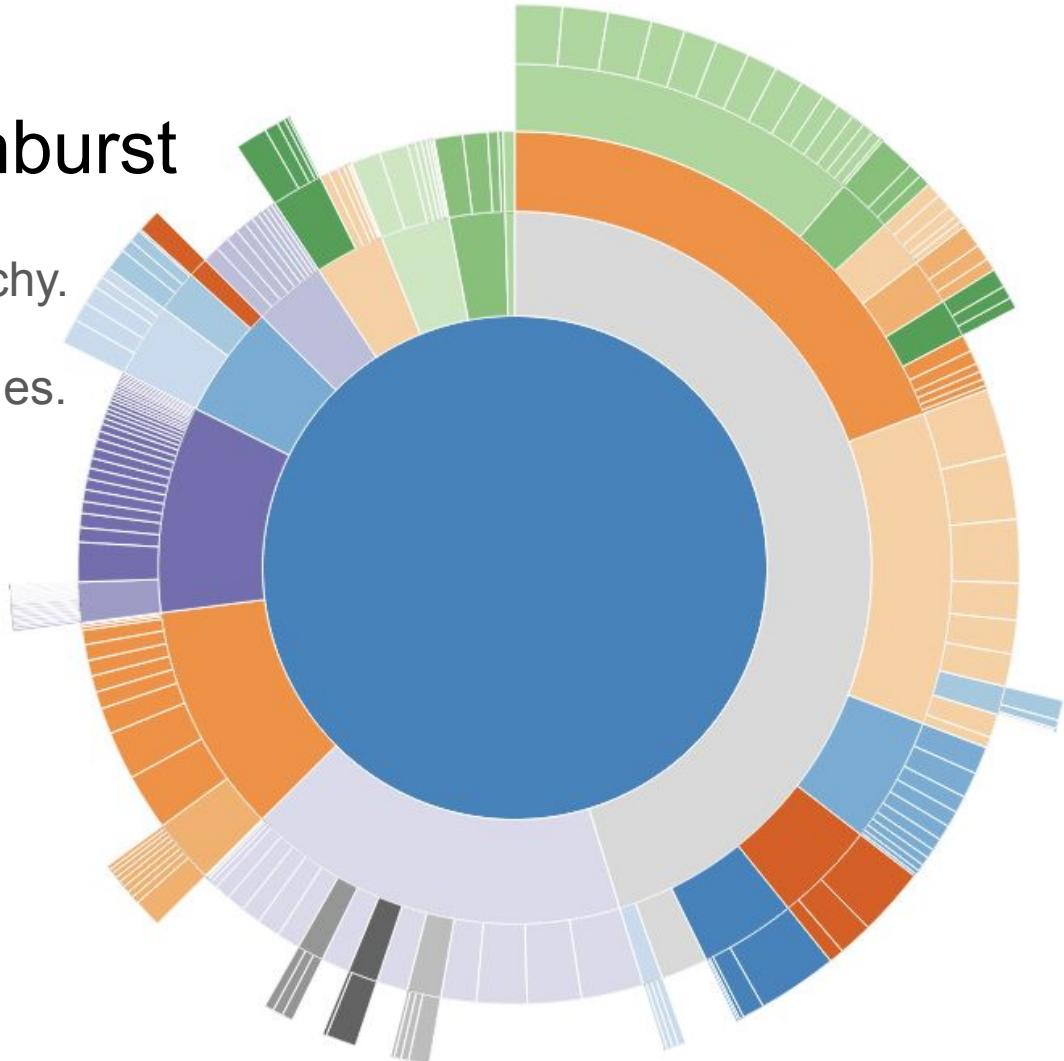
Hierarchical data

Set-typed data

Hierarchical data -- Sunburst

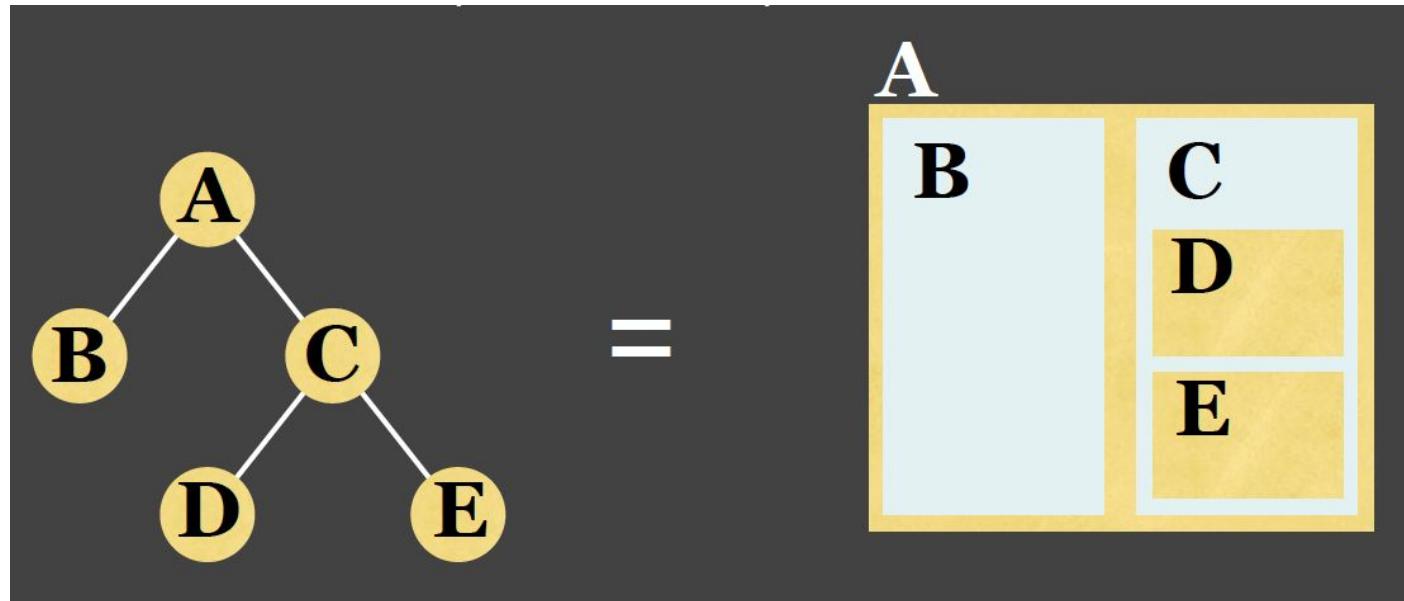
Using adjacency to represent hierarchy.

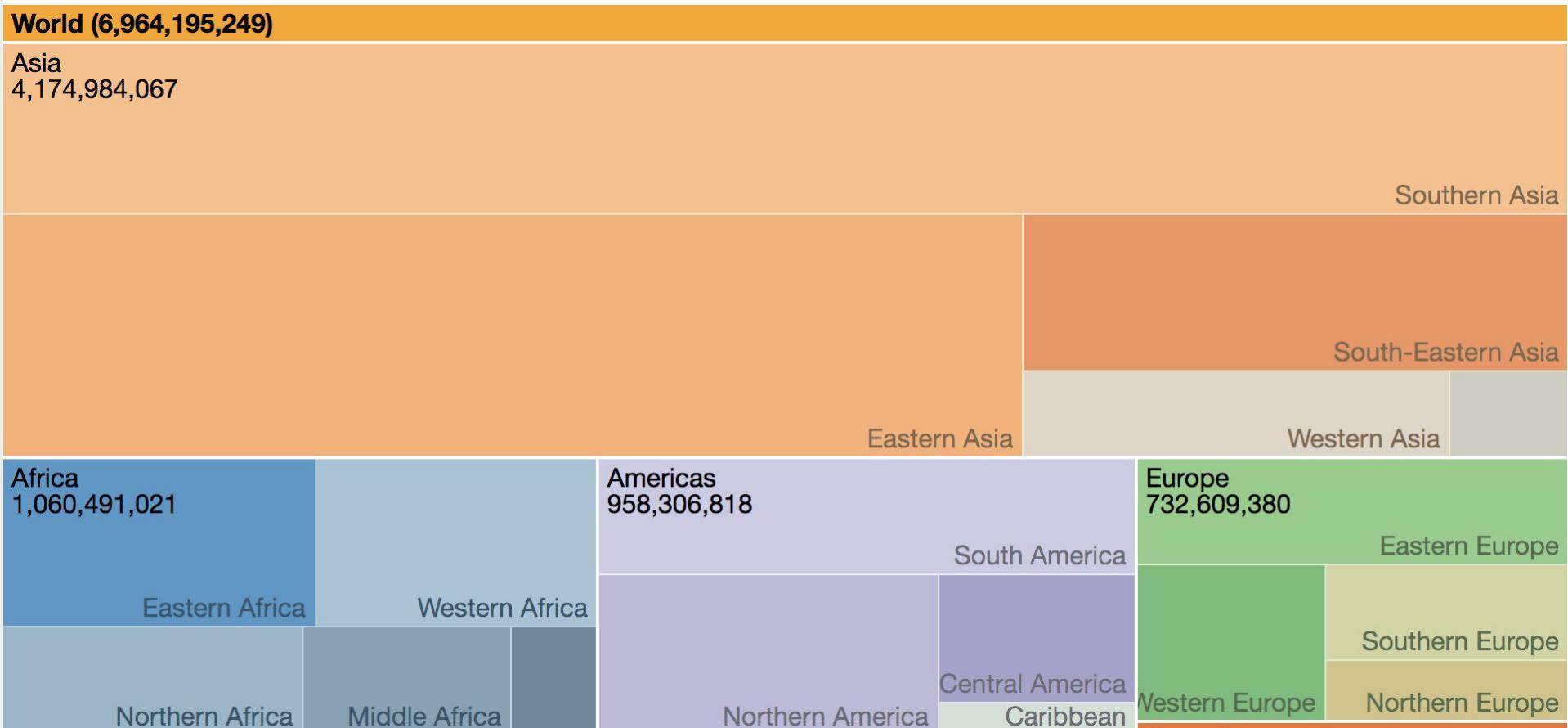
A length encoding for the size of nodes.



Hierarchical data -- Treemaps

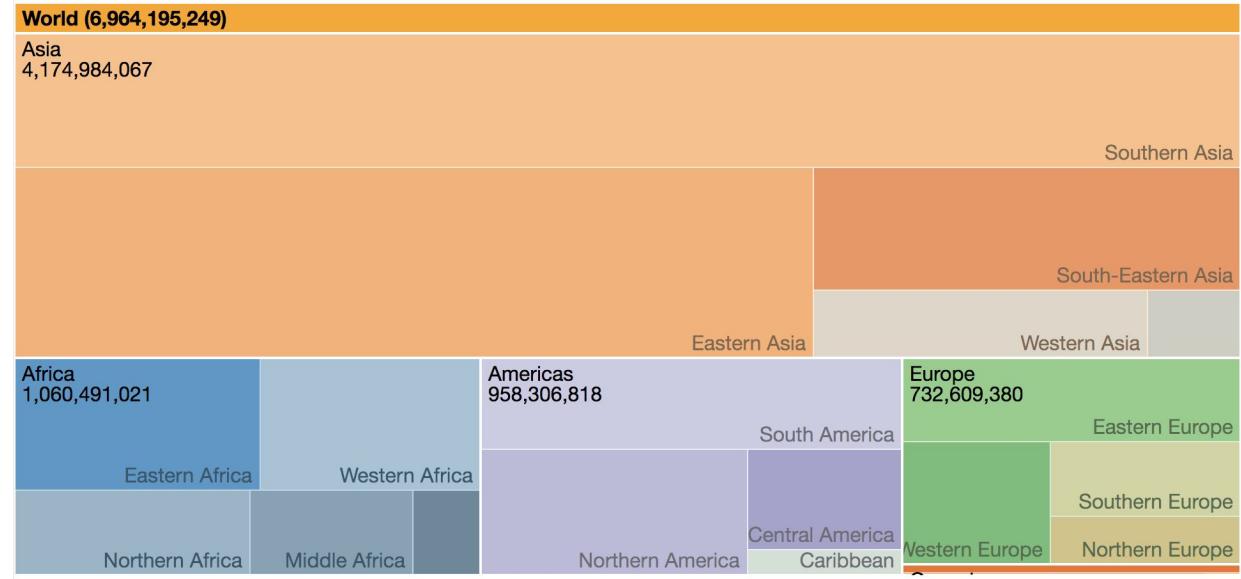
Using **containment** to represent hierarchy;





Hierarchical data -- Treemaps

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



Visualizing different 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.



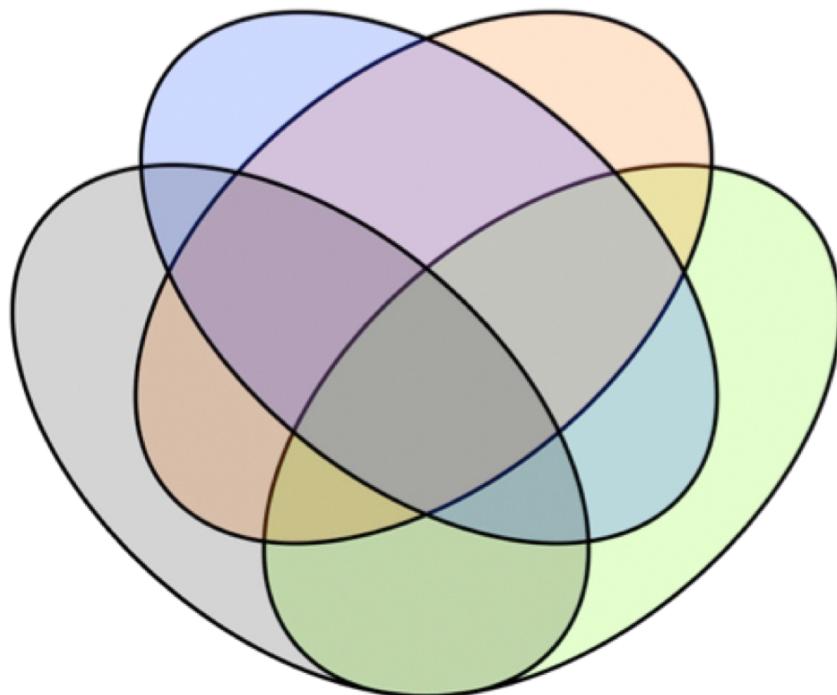
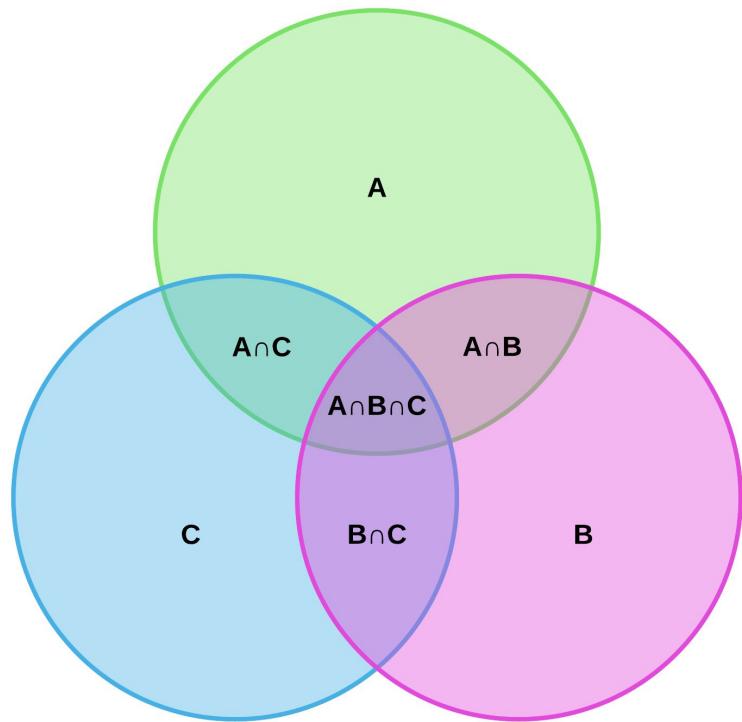
Element

Set



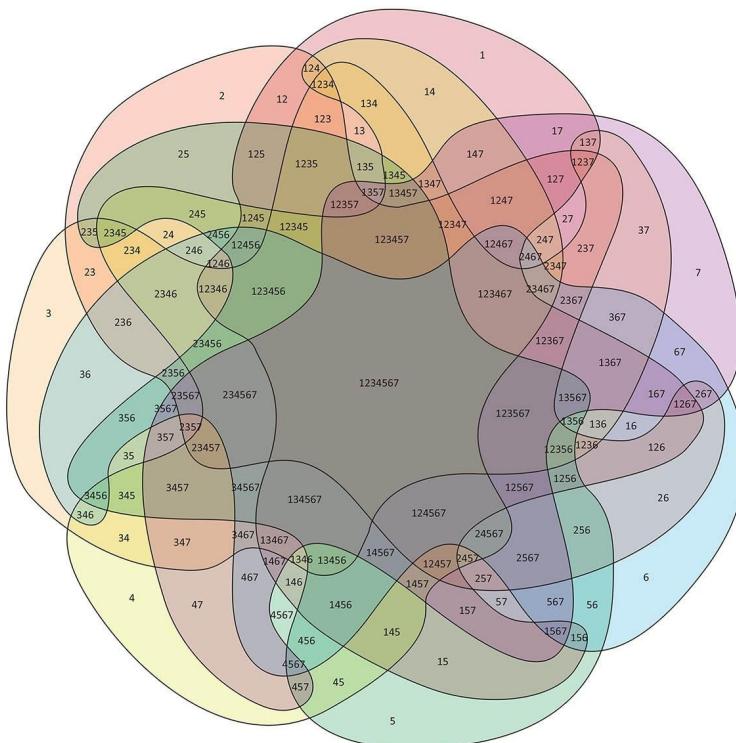
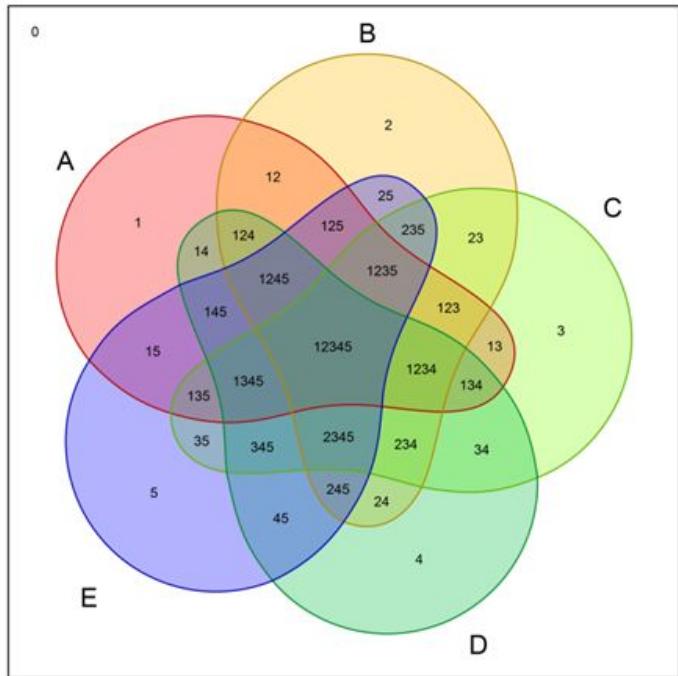
Venn diagrams

Show all possible set relations.



Venn diagrams

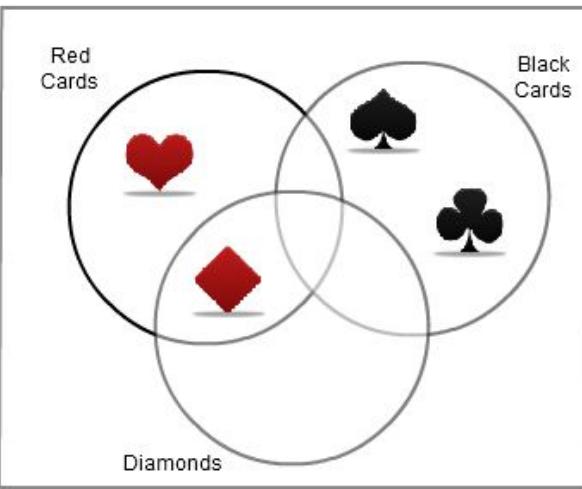
Get messy fast.



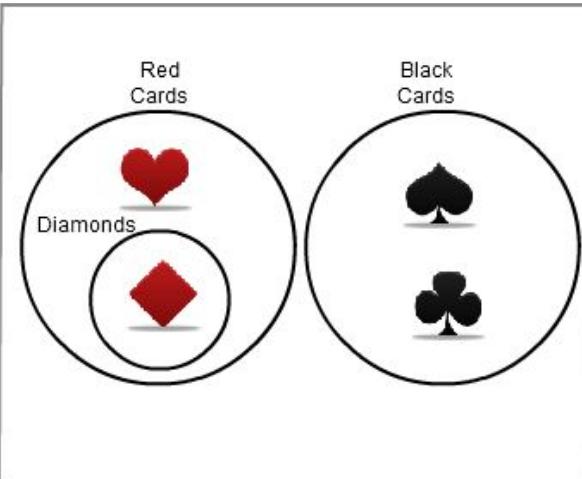
Euler diagrams

Only show existing set relations.

V
E
N
N



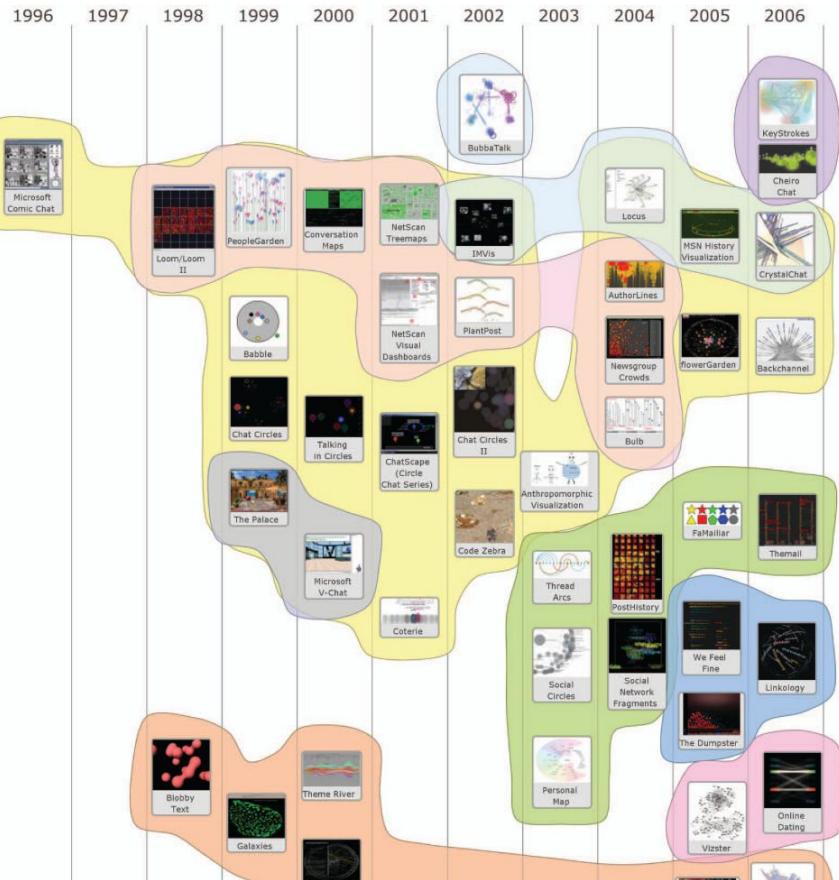
E
U
L
E
R

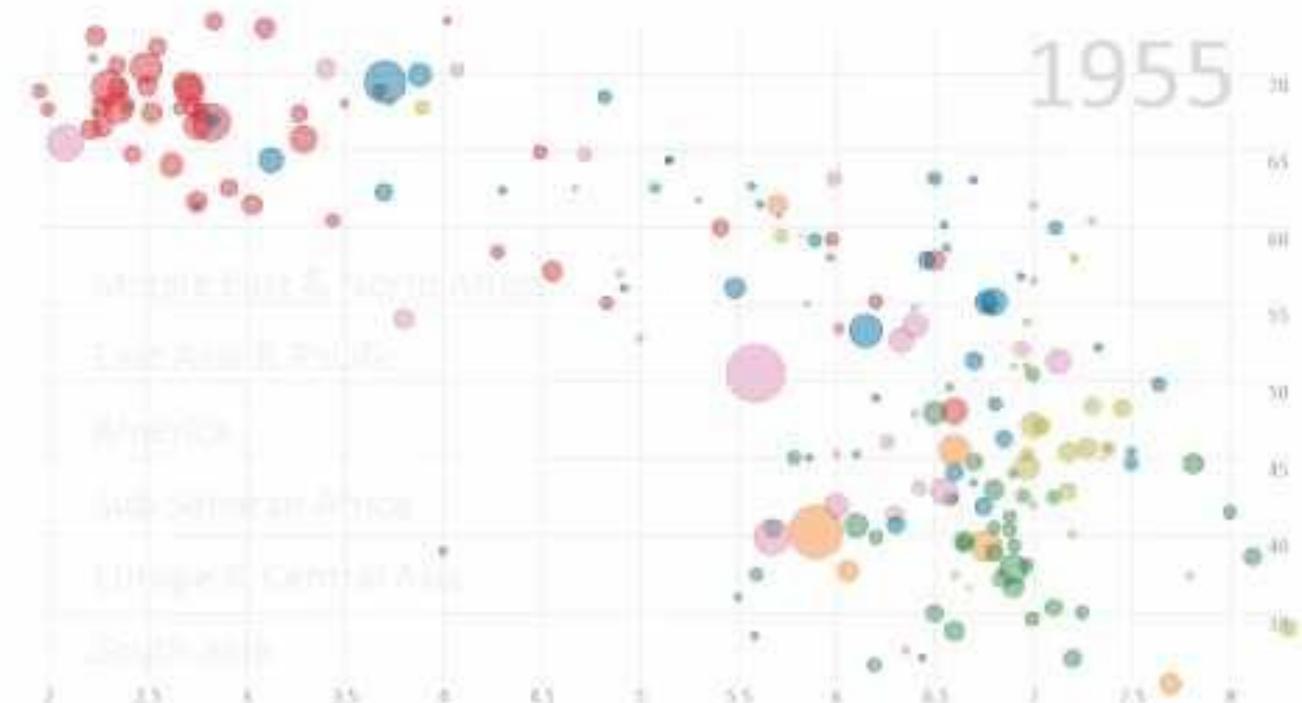


Overlays -- Bubble Sets

Overlay a set membership relation on top of a primary data relation.

Use line contours.

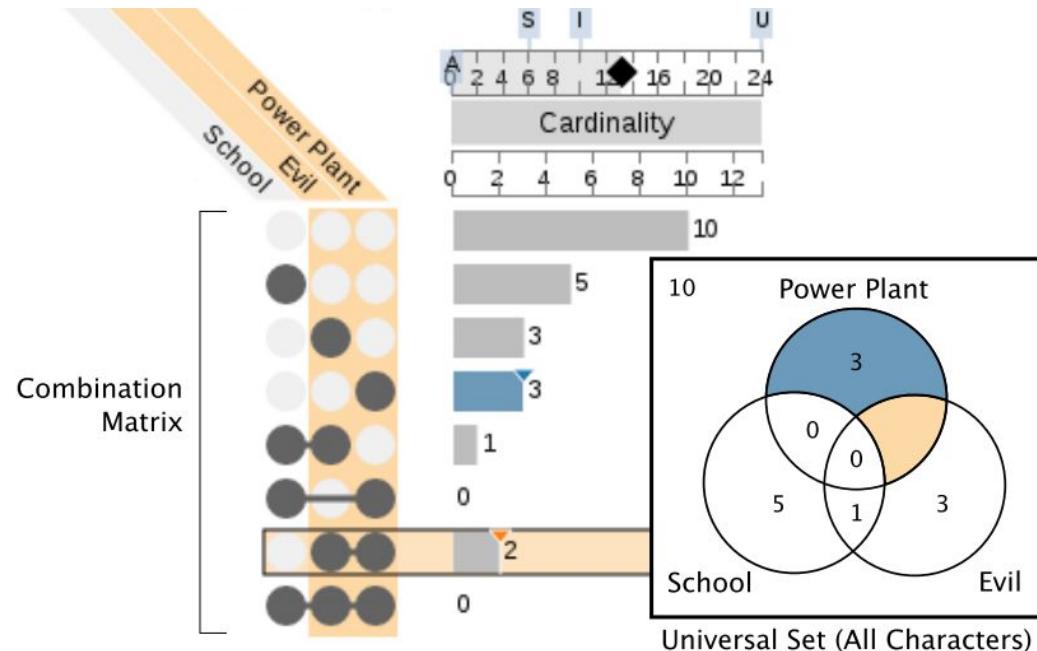




Matrix-based method -- UpSet

Row: set relation

Column: set





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>

Information seeking mantra - Interacting with visualizations

Overview first, zoom and filter, details on demand

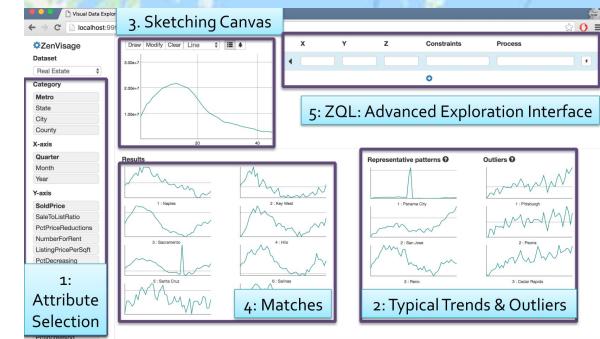
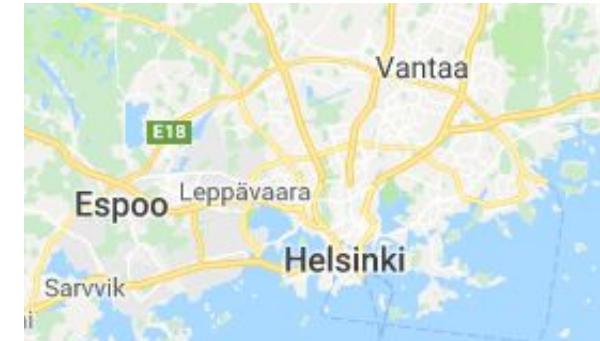
Overview first, zoom and filter, details on demand

Overview first, zoom and filter, details on demand



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

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

Purposes of visualization

To help

Make a decision

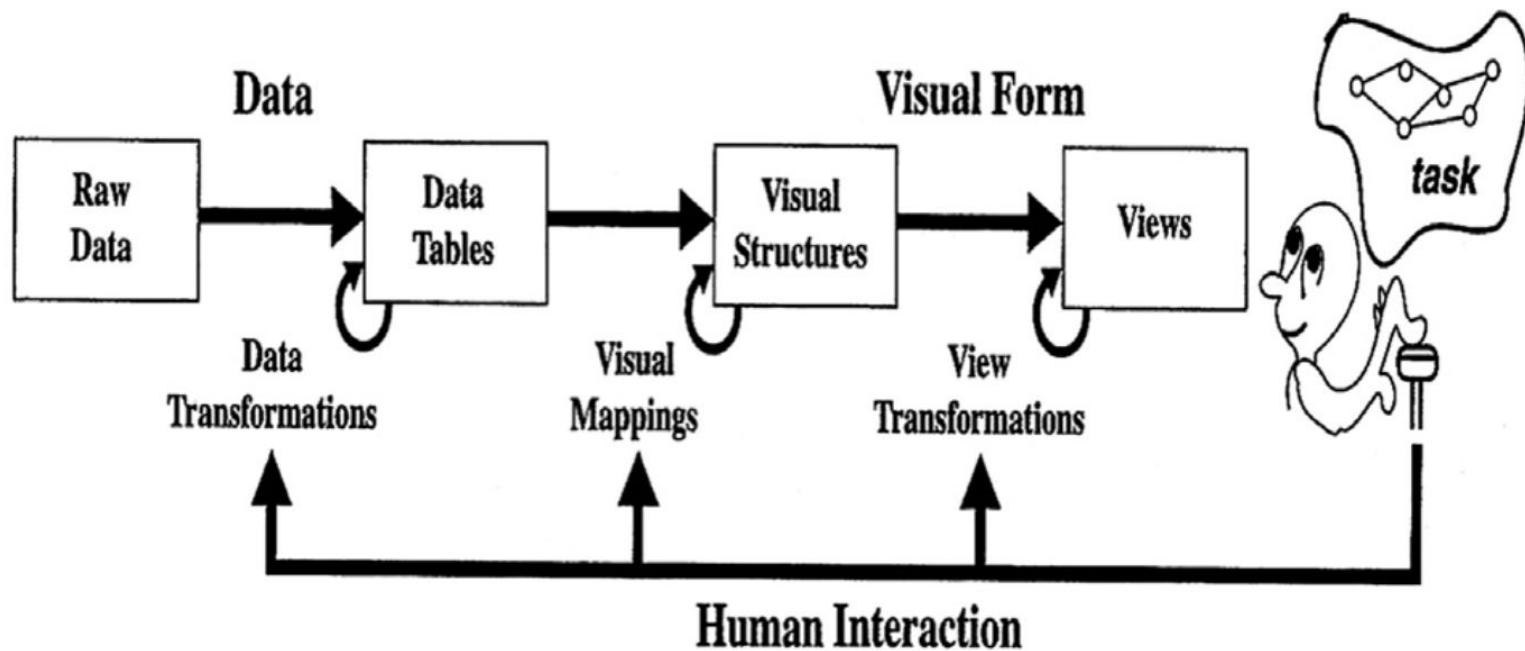
Tell a story

Reasoning

Discover knowledge

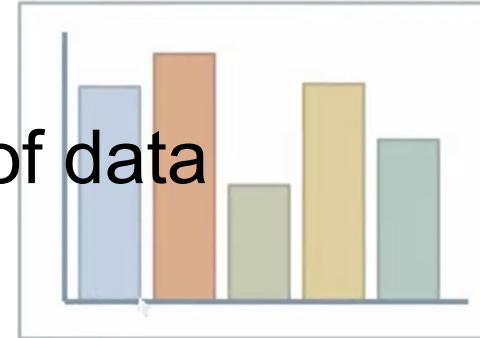
.....

Recap -- Visualization framework

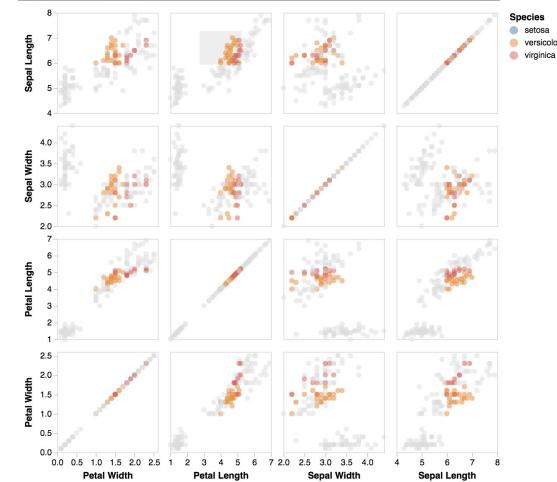


Recap -- Visualizing different types of data

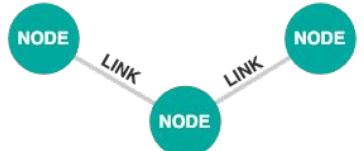
Two-dimensional data



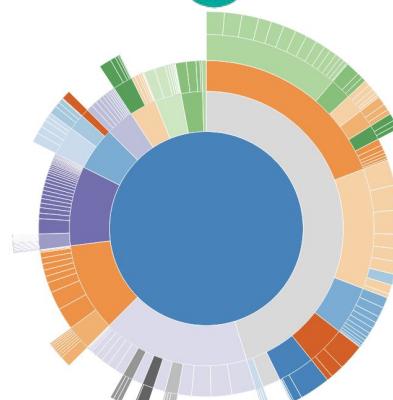
Multi-dimensional data



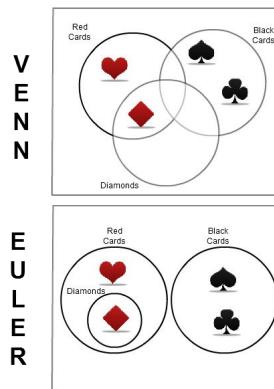
Graph



Hierarchical data



Set-typed data



Recap -- Information seeking mantra

Overview first, zoom and filter, details on demand

DATA15003: Interactive Data Visualization

Human perception

Visualization design

Visualizing **multi-dimensional** data

Visualizing **graph & set-typed** data

Interacting with visualizations

Introduction to **visualization tools**, such as `plotly` and `d3.js`

Applied visualizations for

information retrieval,

augmented and virtual reality,

large public multi-touch displays.

COVER COLOUR
Browse Books by Colour

NUMBER 1
Browse Books

AUTHORS
Browse Books by Author Name

Lens

Pupil

Iris

Cornea

Muscles to move eye

Muscles to adjust lens

Blind spot

Fovea

Retina

Optic nerve to brain

TIME PUBLISHED

CONTENT TIME

TIMELINES
Compare Publication Year and Content Time

Son of the sword

by Lee J. Arden - 2001

Times: 1889-1745

Pages: 324

Keywords: Fiction, Mystery, Detective, Thriller, Suspense, Mystery, Thriller

In this earnest mix of history and romance, in the tradition of Twain's

Connecticut Yankee, a contemporary American travels back in time, here

to 1713 Scotland, where Jacobite rebels are in deep trouble. Also, his story

tells to generate a plot twist unlike anything you've seen before.

England and Scotland have right off and on for centuries, with the English

winning most of the time, but the English are getting

victories. After yet another Scottish defeat, the Bonnie Prince holds

up a great sword and calls upon it to bring me a hero, a Highlander!

The Bohemian bookshelf, Thudt et al., 2012.

Study materials

Colin Ware. **Information Visualization: Perception for Design**, 3rd Edition. Morgan Kaufmann, 2012. [ebook available] (Lecture 3&4)

Edward Tufte. **The Visual Display of Quantitative Information**, 2nd Edition. Graphics Press, 2001. (Lecture 6)

Tamara Munzner. **Visualization Analysis and Design**. CRC Press, 2014. (Lecture 5)

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

Data visualization on Coursera

<https://www.coursera.org/learn/datavisualization/home/welcome>

Tools to create visualization

Tableau Public -- Free data visualization software

<https://public.tableau.com/en-us/s/>

Many eyes -- IBM's free online data visualization tool

<http://www.boostlabs.com/ibms-many-eyes-online-data-visualization-tool/>

Plotly <https://plot.ly/>

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

D3.js <https://d3js.org/>

Some visualization-related talks

Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats.

<https://youtu.be/jbkSRLYSojo>

The genius of the London tube map.

https://www.ted.com/talks/michael_bierut_the_genius_of_the_london_tube_map

The beauty of diagrams.

<https://www.bbc.co.uk/programmes/b00w5675> or <https://www.youtube.com/playlist?list=PL2364C9DD05C26465>

The beauty of data visualization.

<https://youtu.be/5Zg-C8AAIGg>

Some human perception & cognition related videos

National Geographic: Test Your Brain Episode 1 - Pay Attention.

<https://www.dailymotion.com/video/xq1p3e>

National Geographic: Test Your Brain Episode 2 - Perception.

<https://www.dailymotion.com/video/xq1rf1>

National Geographic: Test Your Brain Episode 3 - Memory.

<https://www.dailymotion.com/video/x1jpdsw>

Very interesting!

Thank you! Q&A

Chen He

chen.he@helsinki.fi