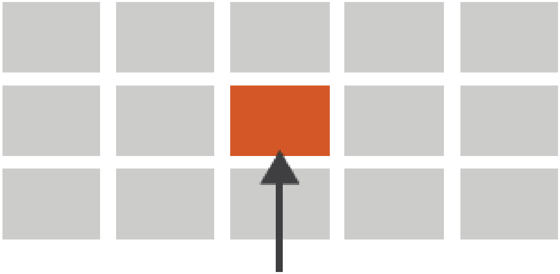
Introduction to Visualization

part 2

Noeska Smith, Jan Byška *et al.*,
UiB Dept. of Informatics,
2017-08-28

Various Types of Data

table data



spatial data

network data

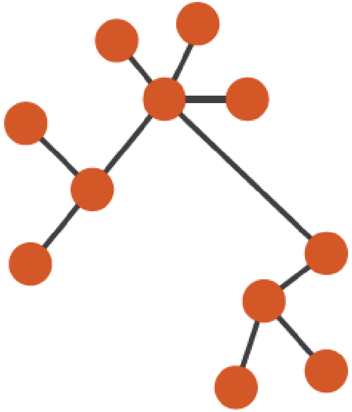


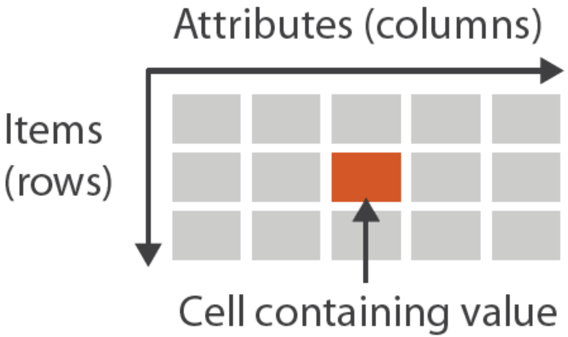
Table Data

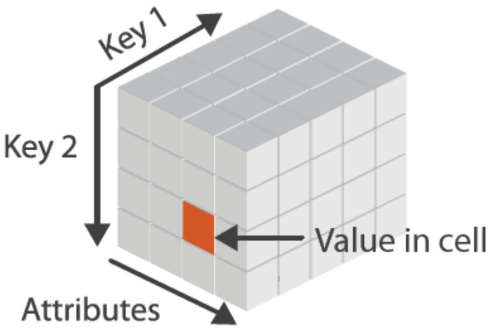
key

- unique independent attribute
  - row index

values

- dependent attribute, value of a cell
- quantitative vs. categorical
  - animal weight vs. animal species





Scatterplot

no keys, only values

– 2 quantitative attributes

mark: points

channels

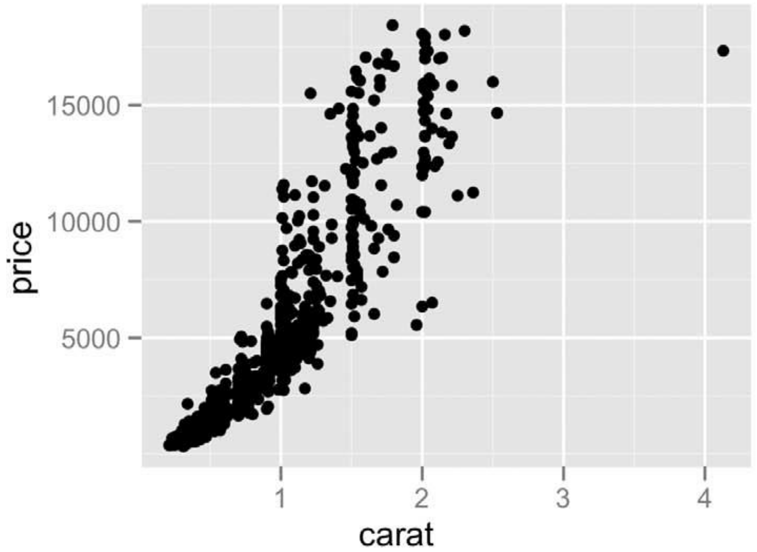
– horizontal + vertical position

tasks

– find trends, outliers, distribution, correlation, clusters

scalability

– hundreds of items



A layered grammar of graphics. Wickham. J. Comp. Graph. Stat., 2010.

Bar Chart

one key, one value

– 1 categorical attribute, 1 quantitative attribute

mark: lines

channels

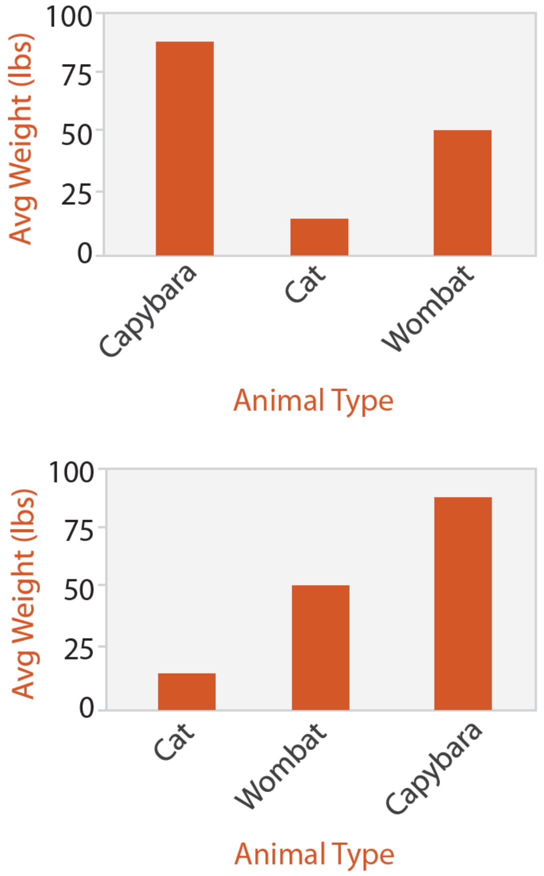
– length to express quantitative values
– spatial regions: one per mark
  - ordered by label (alphabetical),
  - ordered by length attribute (data-driven)

task

– compare, lookup values

scalability

– dozens to hundreds of levels for key attribute



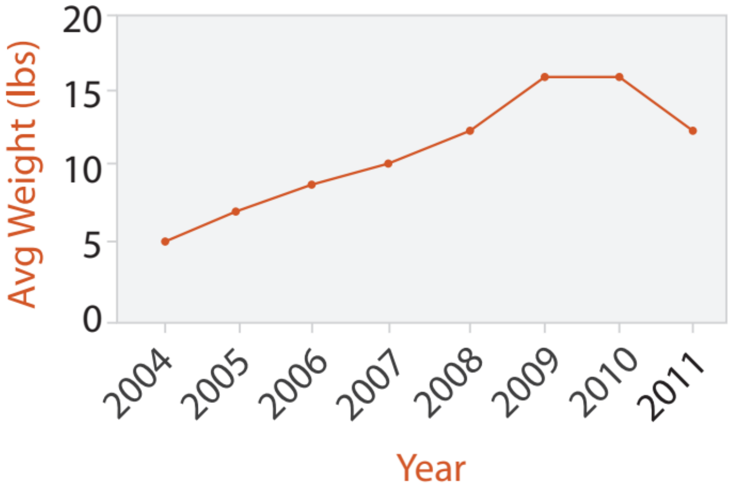
Line Chart / Dot Plot

one key, one value

- 2 quantitative attributes

mark: points

- plus line connection marks
  - explicitly showing relationship between items



channels

- length (position) to express quantitative values
- separated and ordered by key attribute

task

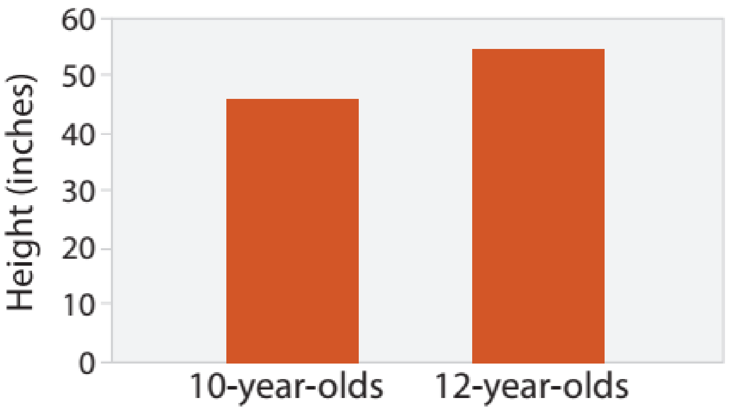
- find trend

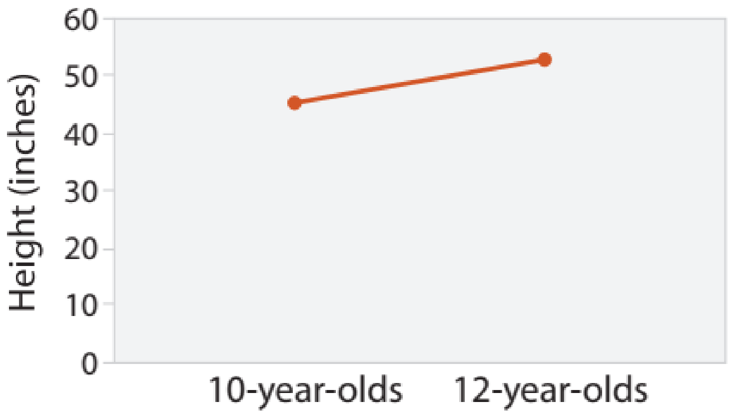
Choosing Bar or Line Chart

depends on type of key attribute

– bar charts if categorical

– line charts if ordered (quantitative)

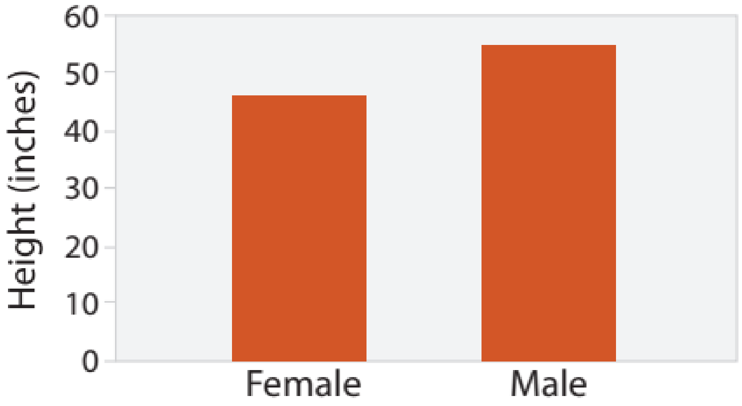


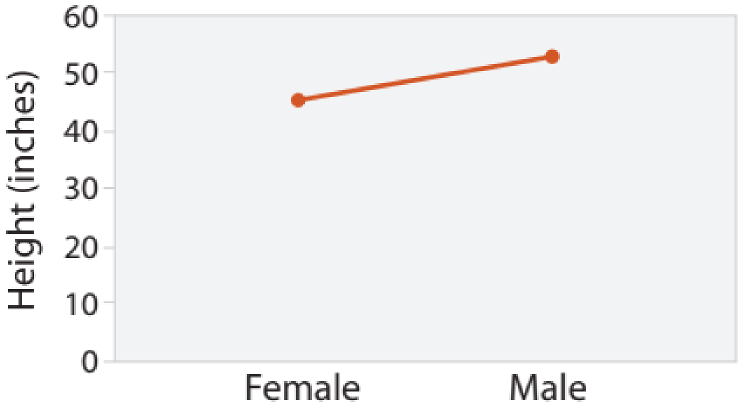


Choosing Bar or Line Chart

do not use line charts for categorical key attributes

– implication of trend so strong that it overrides semantics!
  - "The more male a person is, the taller he/she is."





Stacked Bar Chart

data: 2 categorical attributes, 1 quantitative attribute

mark: vertical stack of line marks

channels

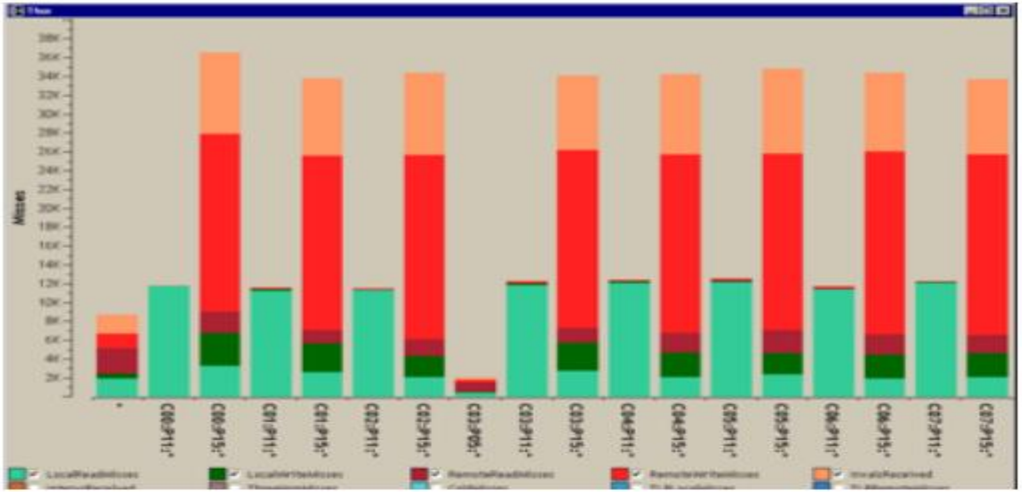
- length
- color hue
- spatial regions
  - one per stacked bar

task

- part-to-whole relationship

scalability

- several to one dozen levels for stacked attribute



Using Visualization to Understand the Behavior of Computer Systems. Bosch. Ph.D. thesis, Stanford Computer Science, 2001.

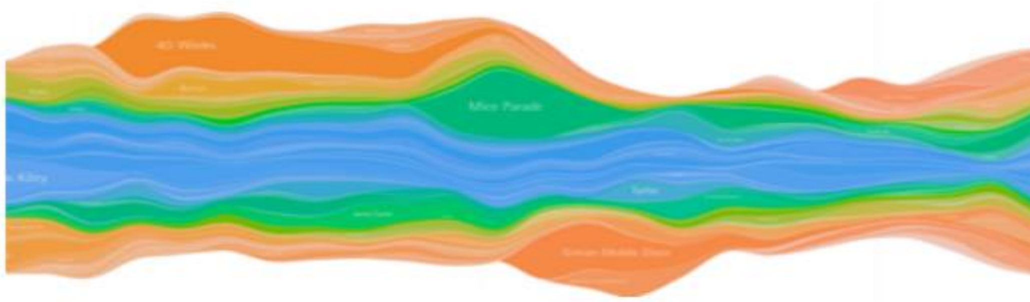
Streamgraph

generalized stacked graph

- emphasizing horizontal continuity vs. vertical items

data

- 1 categorical key attribute (artist)
- 1 ordered key attribute (time)
- 1 quantitative value attribute (counts)



Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE TVCG, 2008.

derived data

- geometry: layers -> height encodes counts
- 1 quantitative attribute (layer ordering)

scalability

- hundreds of time keys
- dozens to hundreds of artist keys
  - more than stacked bars, since most layers don't extend across whole chart

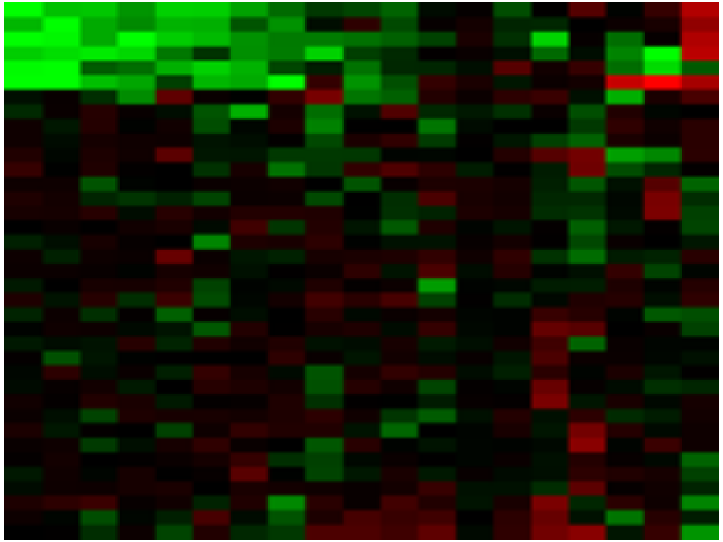
Heatmap

two keys, one value

- 2 categorical attributes (gene, experimental condition)
- 1 quantitative attributes (expression levels)

marks: area

- separate and align in 2D matrix
  - indexed (ordered) by 2 categorical attributes



channels

- color by quantitative attribute
  - ordered vs. diverging color map

task

- find clusters, outliers

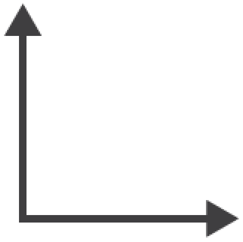
scalability

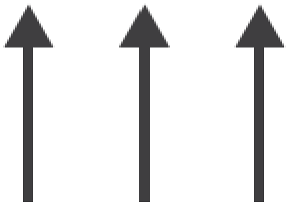
- 1K categorical levels, 1M items; only ~10 quantitative attribute levels

Multiple Axis

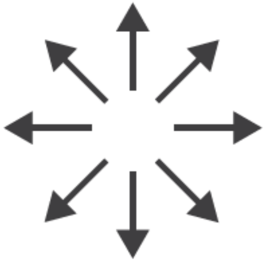
Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

rectilinear





parallel



radial

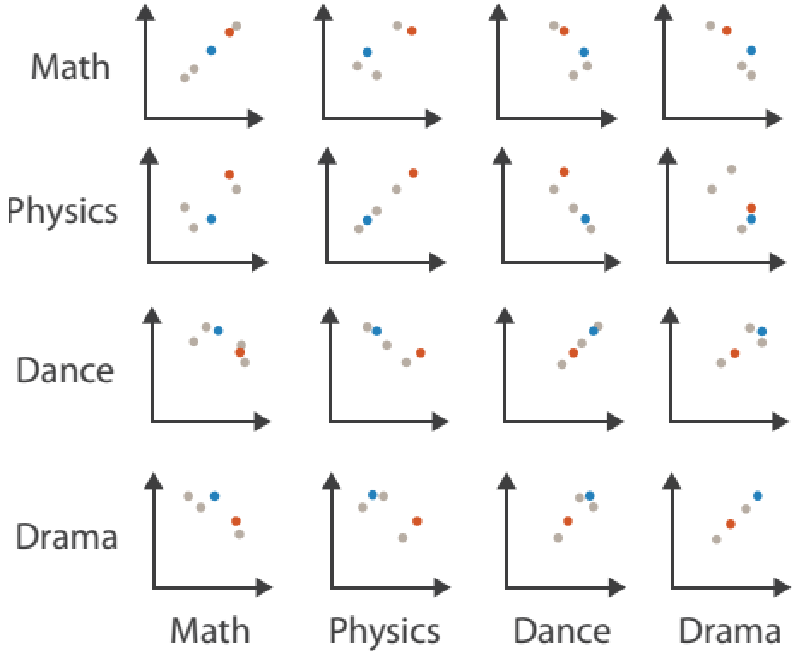
Scatterplot Matrix

rectilinear axes, point mark

- all possible pairs of axes

scalability

- one dozen attributes
- dozens to hundreds of items



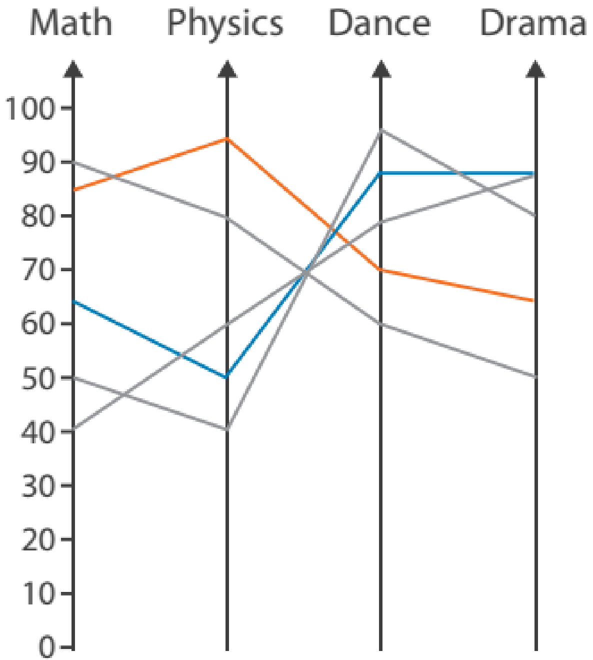
Parallel Coordinates

parallel axes, jagged line representing item

- parallel axes, item as point
  - axis ordering is major challenge

scalability

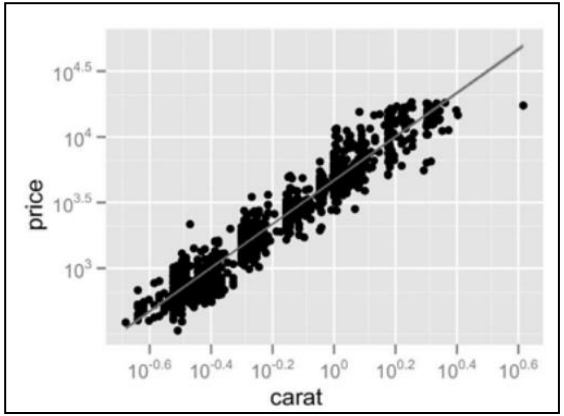
- dozens of attributes
- hundreds of items



Task: Correlation

scatterplot matrix

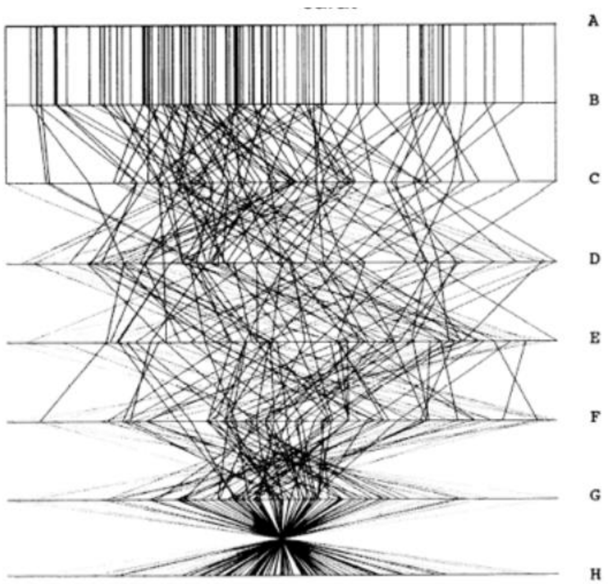
- positive correlation
  - diagonal low-to-high
- negative correlation
  - diagonal high-to-low
- uncorrelated
  - scattered points



A layered grammar of graphics. Wickham. J. Comp. Graph. Stat. 19:1, 2010.

parallel coordinates

- positive correlation
  - parallel line segments
- negative correlation
  - all segments cross at halfway point
- uncorrelated
  - scattered crossings



Hyper dimensional Data Analysis Using Parallel Coordinates. Wegman. JASA, 1990.

Pie chart, Polar Area Chart

data

- 1 categorical key attribute
- 1 quantitative value attribute

pie chart

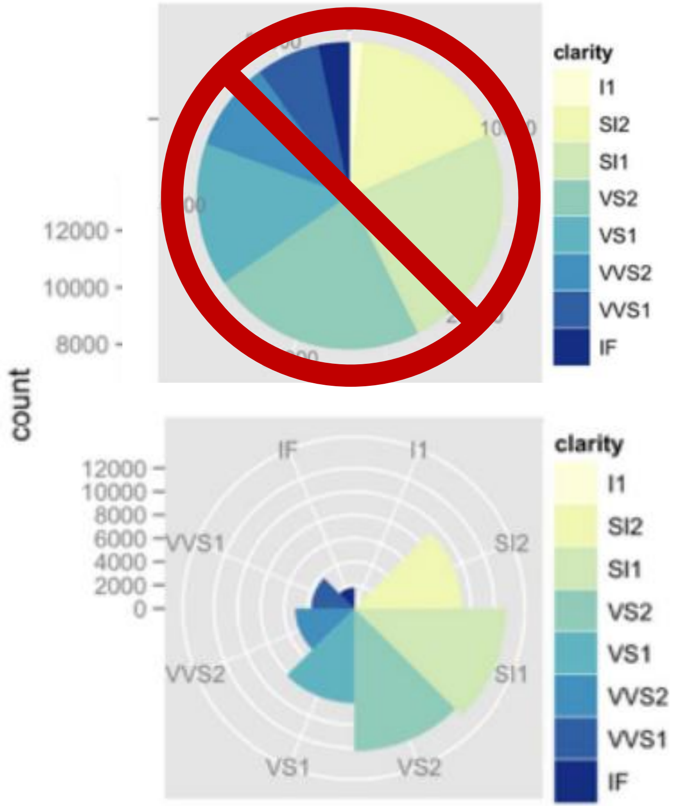
- area marks with angle channel
- accuracy: angle/area much less accurate

polar area chart

- area marks with length channel
- more direct analog to bar charts

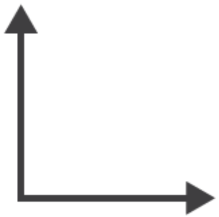
task

- part-to-whole judgements

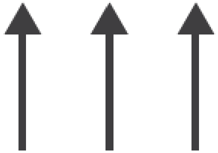


Orientation Limitations

rectilinear: scalability w.r.t. #axes

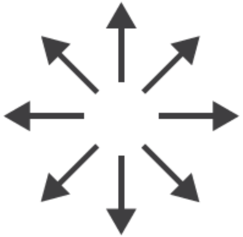


- 2 axes best
- 3 problematic
- 4+ impossible



parallel: unfamiliarity, training time

radial: perceptual limits

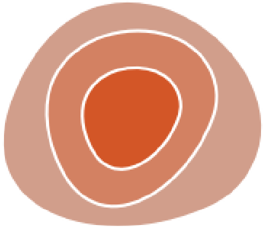


- angles lower precision than lengths
- asymmetry between angle and length

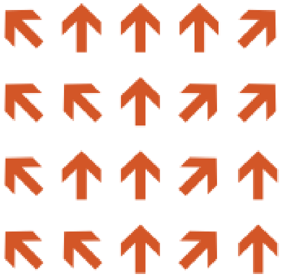
Spatial Data

geometry
- geographical, other derived

spatial fields (one value per cell)
- scalar fields (one value per cell)
  - iso-contours
  - direct volume rendering



- vector and tensor fields (many values per cell)
  - flow glyphs (local)
  - geometric (sparse seeds)
  - textures (dense seeds)
  - features (globally derived)



Choropleth Map

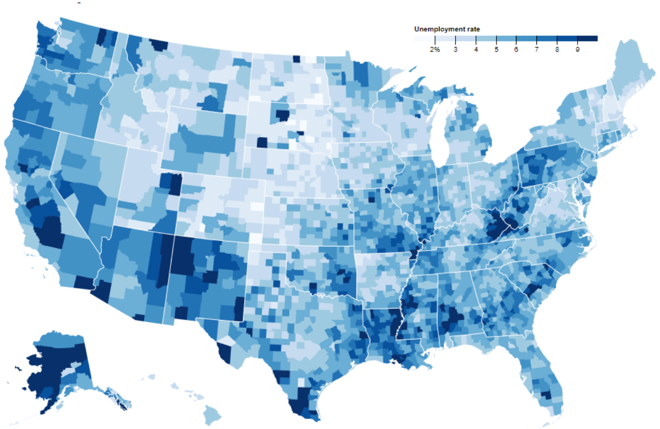
table with 1 quantitative attribute per region

encoding

- geometry for area mark boundaries
- sequentially segmented color map

task

- understanding spatial relationships



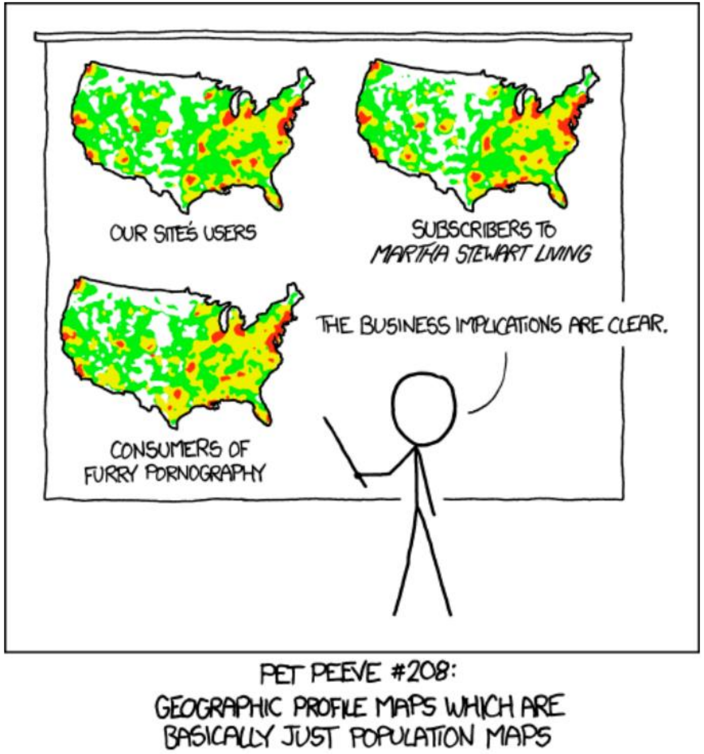
https://bl.ocks.org/mbostock/4060606

Beware: Population Maps Trickiness!

consider when to normalize by population density

general issue

- absolute counts
- relative/normalized data



https://xkcd.com/1138

Topographic Map

1 quantitative attribute per 2D grid cell

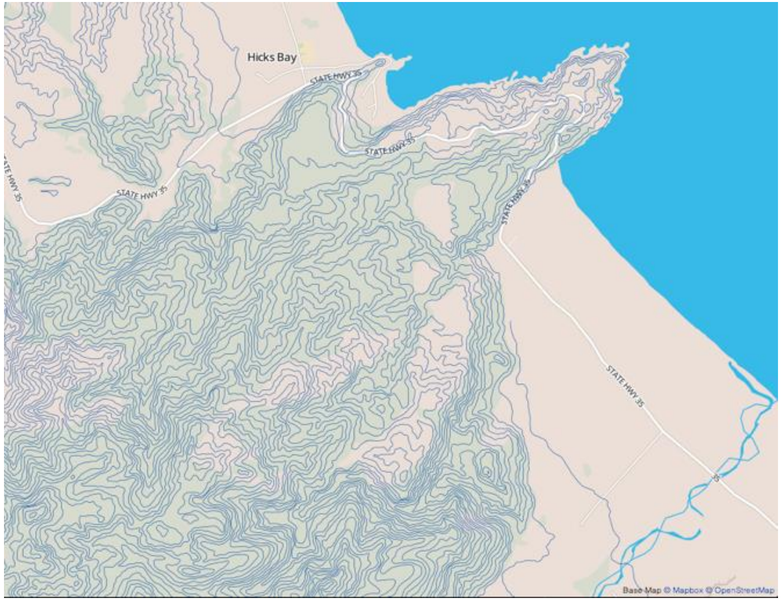
- geographic geometry
- scalar spatial field

task

- shape understanding, spatial relationships

derived data

- isocontours computed for specific levels of scalar values



Land Information New Zealand Data Service

Isosurfaces, Direct Volume Rendering

1 quantitative attribute per 3D grid cell

- scalar spatial field

task

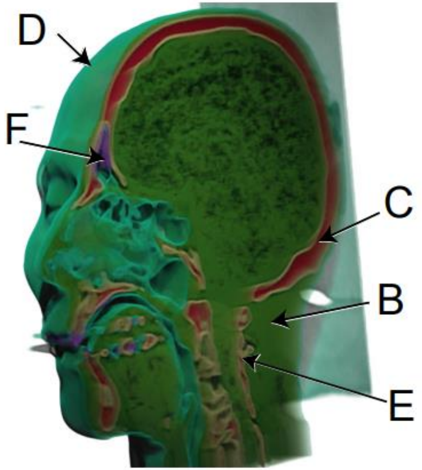
- shape understanding, spatial relationships

derived data

- isocontours computed for specific levels of scalar values

direct volume rendering

- transfer function maps scalar values to color, opacity



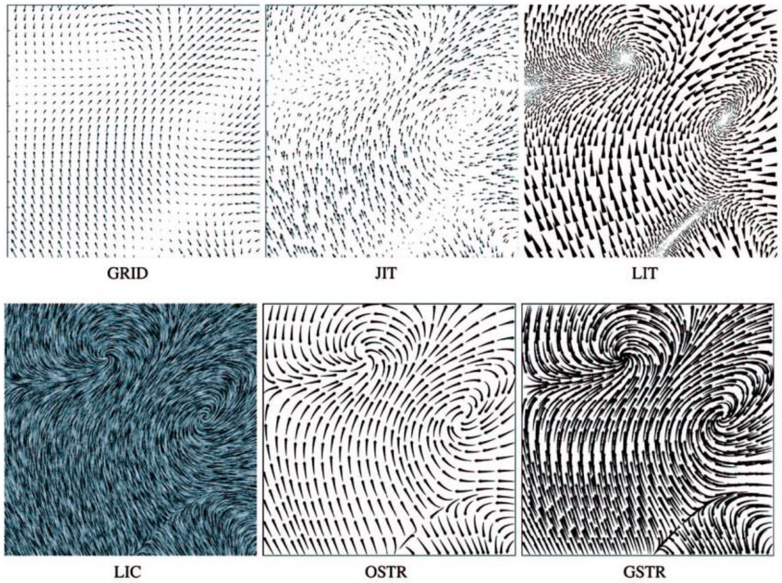
Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. The Visualization Handbook, 2005.

Vector and Tensor Fields

many attributes per cell

idiom families

- flow glyphs
  - purely local
- geometric flow
  - derived data from tracing particle trajectories
  - sparse set of seed points
- texture flow
  - derived data, dense seeds
- feature flow
  - global computation to detect features
  - encoded with one of methods above



Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE TVCG, 2005.

Streamlines

3D vector field

derived data (from field)

– streamlines: trajectory particle will follow

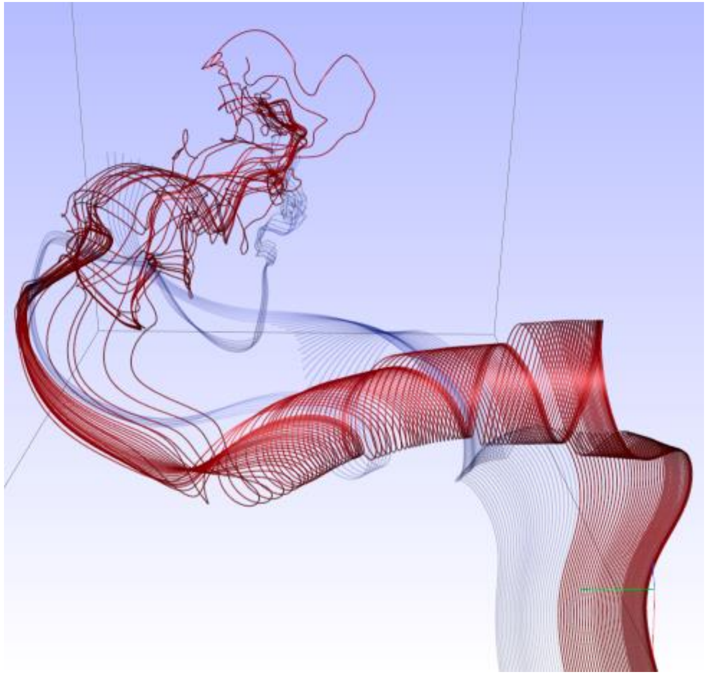
derived data (per streamline)

– curvature, torsion, tortuosity
– signature: complex weighted combination
– compute cluster hierarchy across all signatures
– encode: color and opacity by cluster

tasks

– find features, query shape

- **scalability**
  – millions of samples, hundreds of streamlines

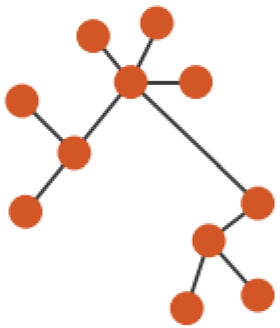


Similarity Measures for Enhancing Interactive Streamline Seeding. McLoughlin,. Jones, Laramee, Malki, Masters, and. Hansen. IEEE TVSG, 2013

Network and Tree Data

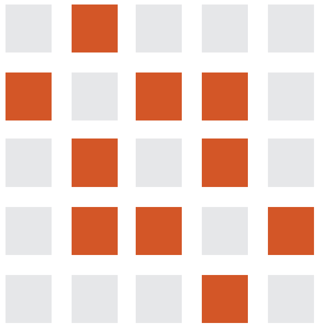
Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES



Force Directed Placement

visual encoding: node-link diagram

- link connection marks, node point marks

algorithm: energy minimization

- analogy: nodes repel, links draw together like springs
- optimization problem: minimize crossings

spatial position: no meaning directly encoded

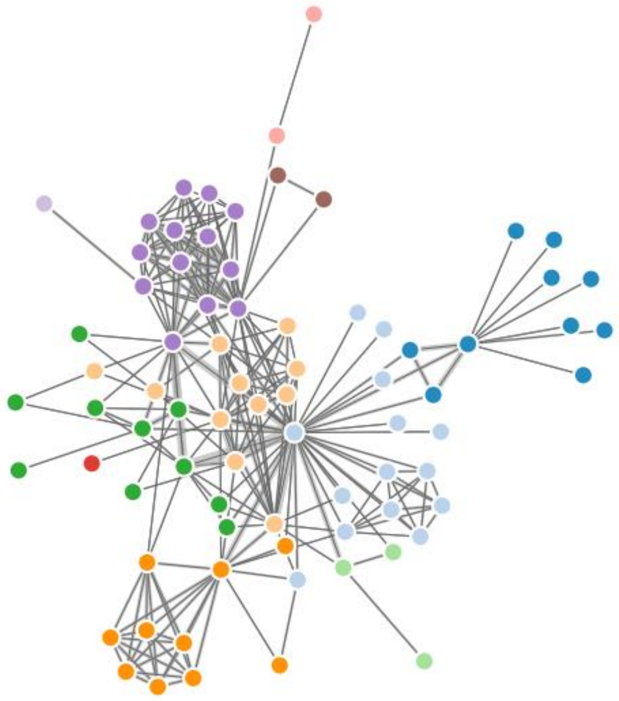
- proximity can be meaningful or arbitrary

tasks

- explore topology; locate paths, clusters

scalability

- node/edge density E < 4N



http://mbostock.github.com/d3/ex/force.html

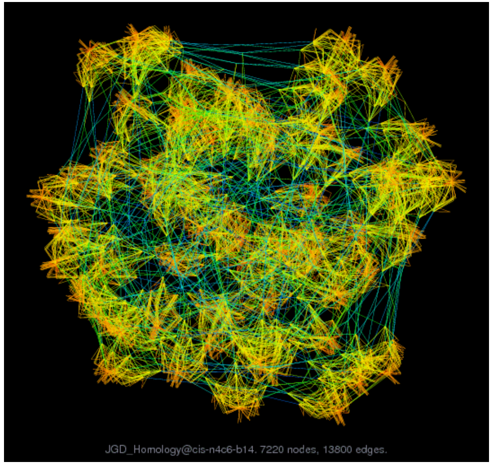
Multilevel Force-Directed Placement

derived data: cluster hierarchy

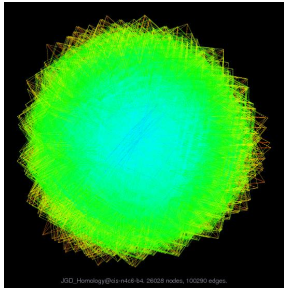
- better algorithm for same encoding technique
- same: fundamental use of space

scalability

- nodes, edges: 1K-10K
- hairball problem still hits eventually



Efficient and high quality force-directed graph drawing. Hu. The Mathematica Journal, 2005.



http://www.research.att.com/yifanhu/GALLERY/GRAPHS/index1.html

Adjacency Matrix View

data: network

- transform into same data/encoding as heatmap

derived data: table from network

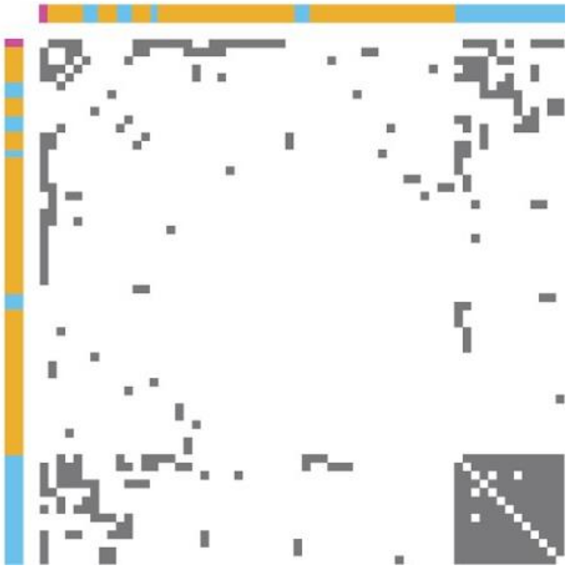
- 1 quantitative attribute
  - weighted edge between nodes
- 2 categorical attributes: node list x 2

visual encoding

- cell shows presence/absence of edge

scalability

- 1K nodes, 1M edges



Points of view: Networks. Gehlenborg and Wong. Nature Methods, 2012

Connection vs. Adjacency Comparison

adjacency matrix strengths

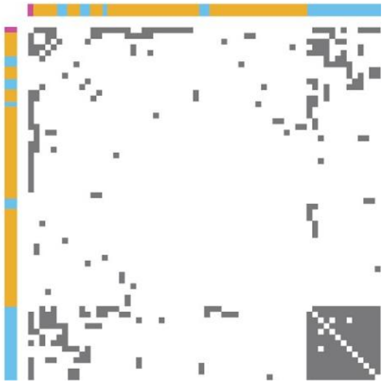
- predictability, scalability, supports reordering
- some topology tasks trainable

node-link diagram strengths

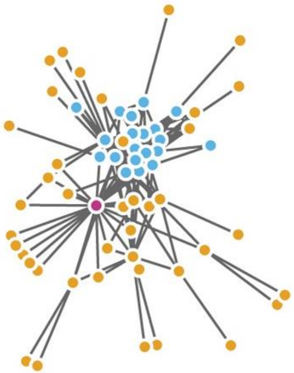
- topology understanding, path tracing
- intuitive, no training needed

empirical study

- node-link best for small networks
- matrix best for large networks
  - if tasks don't involve topological structure!



Points of view: Networks. Gehlenborg and Wong. Nature Methods, 2012



Radial Node-Link Tree

data: tree

encoding

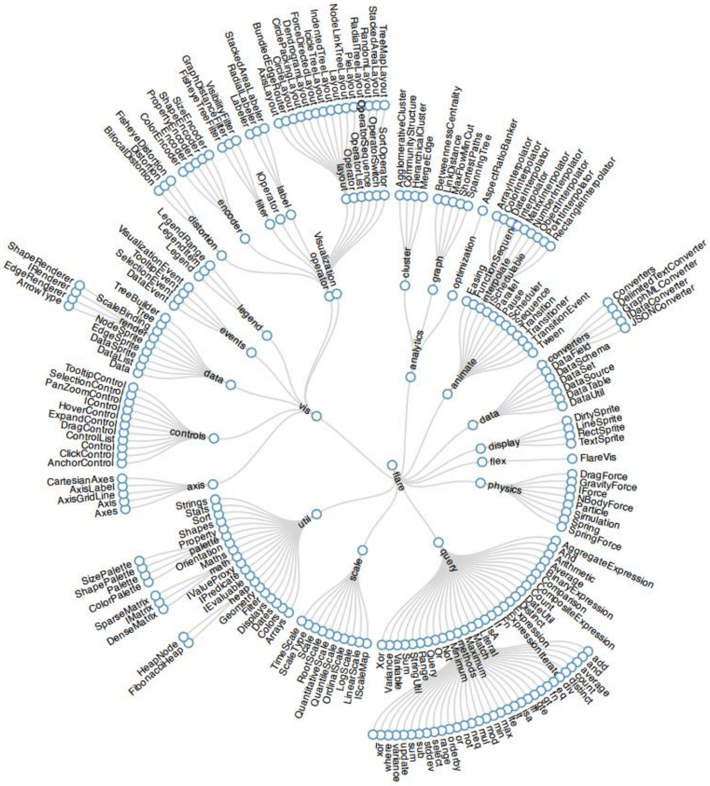
- link connection marks
- point node marks
- radial axis orientation
  - angular proximity: siblings
  - distance from center: depth in tree

tasks

- understanding topology, following paths

scalability

- 1K - 10K nodes



http://mbostock.github.com/d3/ex/tree.html

Treemap

data: tree

- 1 quantitative attribute at leaf nodes

encoding

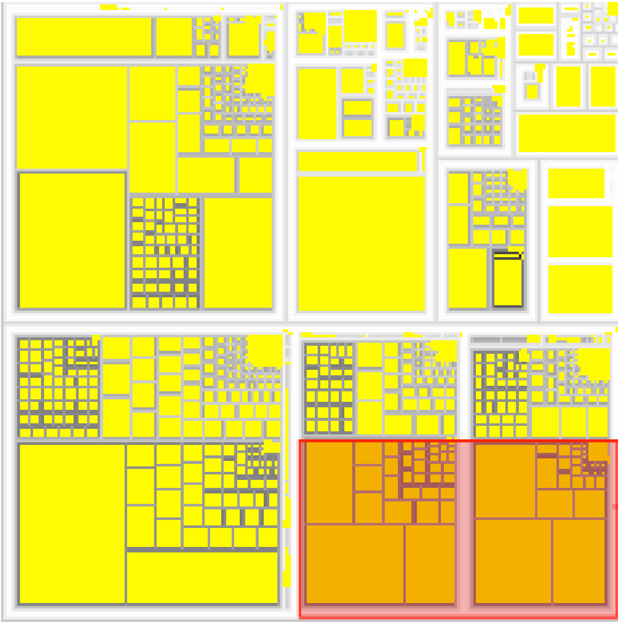
- area containment marks for hierarchical structure
- rectilinear orientation
- size encodes quantitative attribute

tasks

- query attribute at leaf nodes

scalability

- 1M leaf nodes



http://tulip.labri.fr/Documentation/3_7/userHandbook/html/ch06.html

Introduction to Visualization

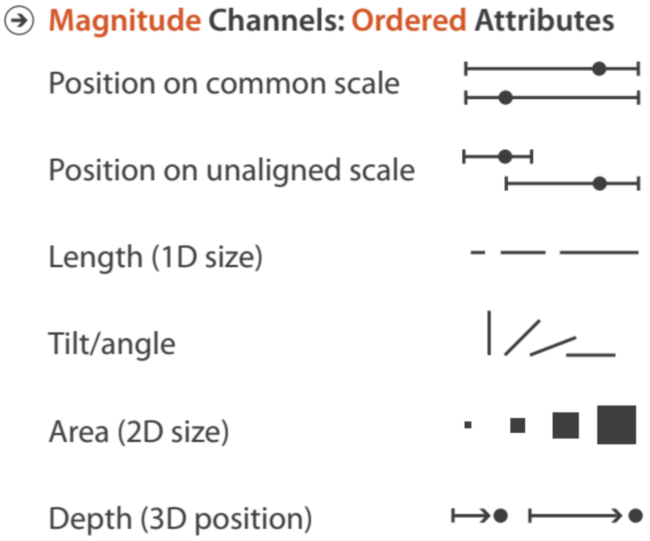
part 3

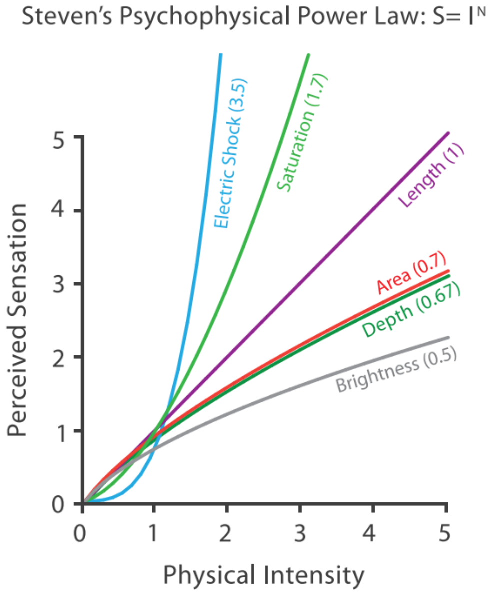
Noeska Smith, Jan Byška *et al.*,
UiB Dept. of Informatics,
2017-08-28

No Unjustified 3D: Power of the Plane

high-ranked spatial position channels:

– planar spatial position **not depth**!

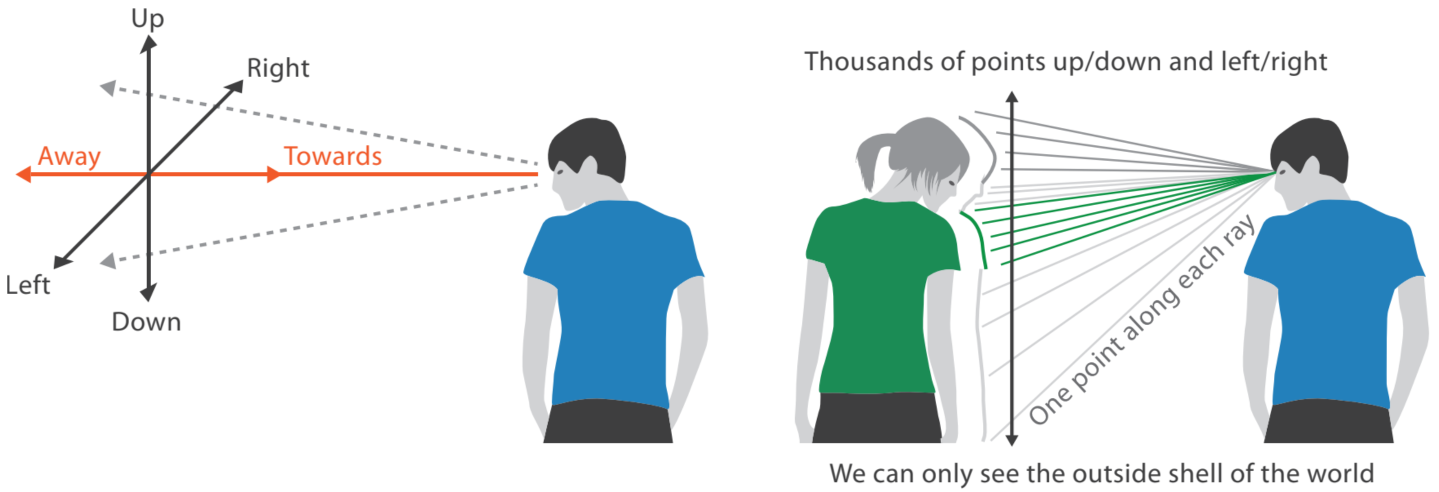




No Unjustified 3D: Power of the Plane

we don't really live in 3D: we see in 2.5D

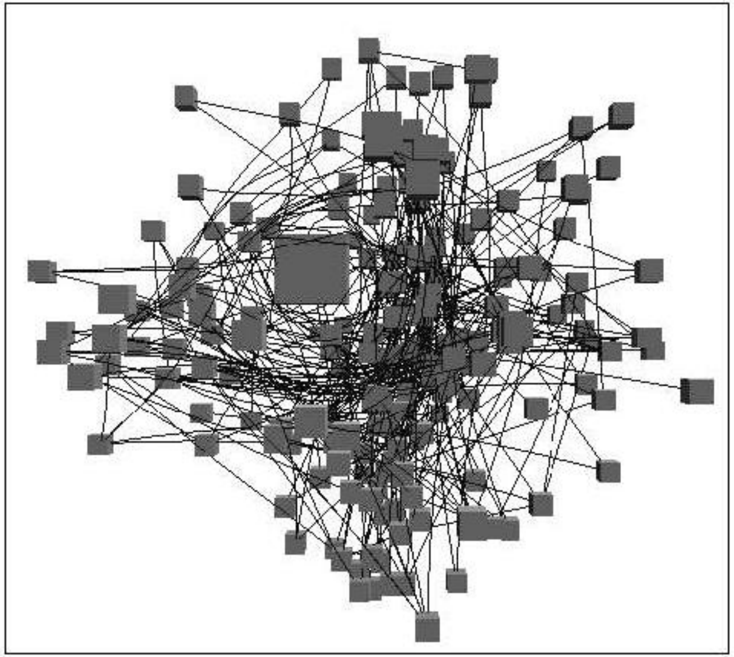
- acquire more info on image plane quickly from eye movements
- acquire more info for depth slower, from head/body motion



No Unjustified 3D: Occlusion

occlusion

interaction complexity

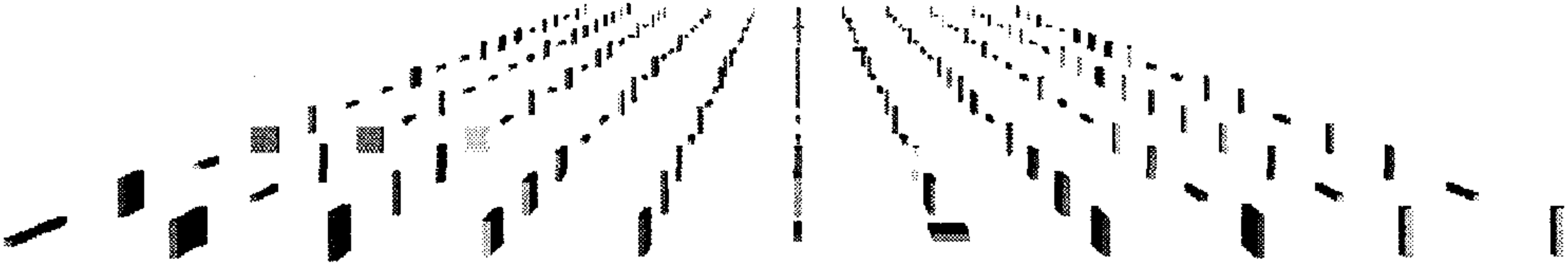


Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis, 1996.

No Unjustified 3D: Distortion

perspective distortion

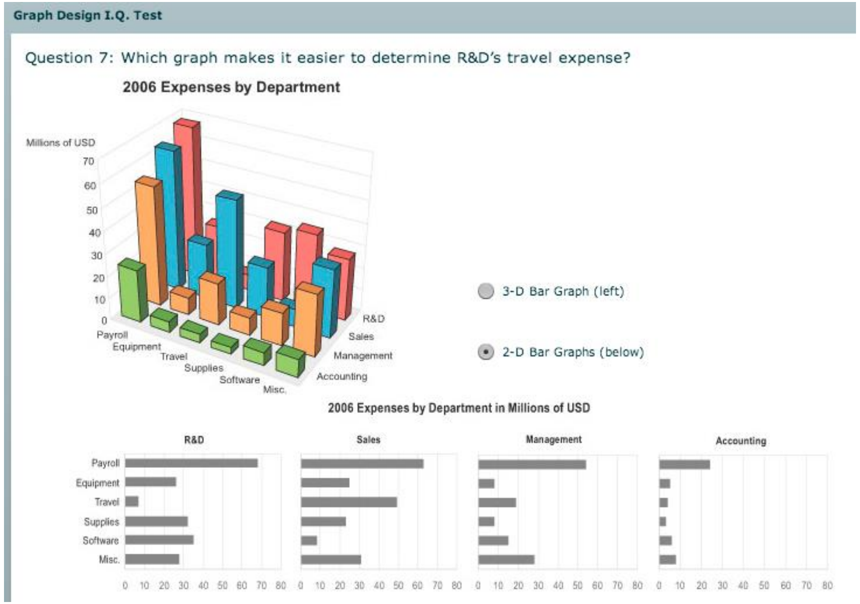
– interferes with all size channel encodings



Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis, 1996

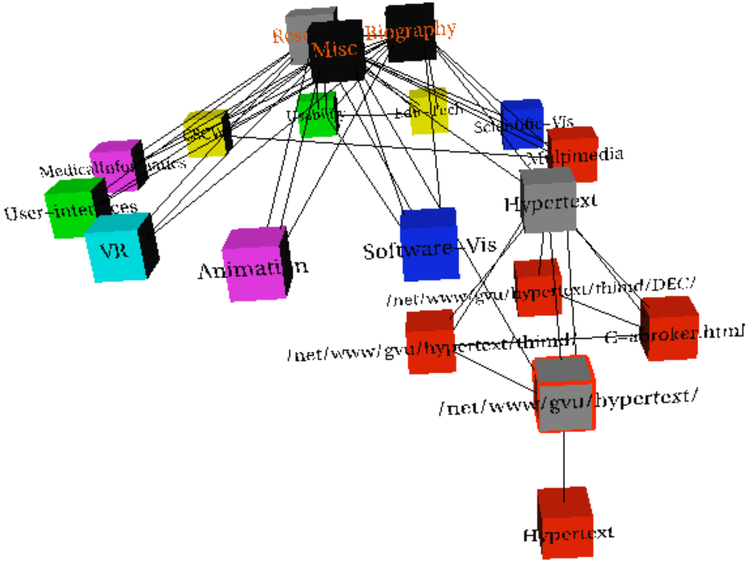
No Unjustified 3D: Examples

3D bars never a good idea!



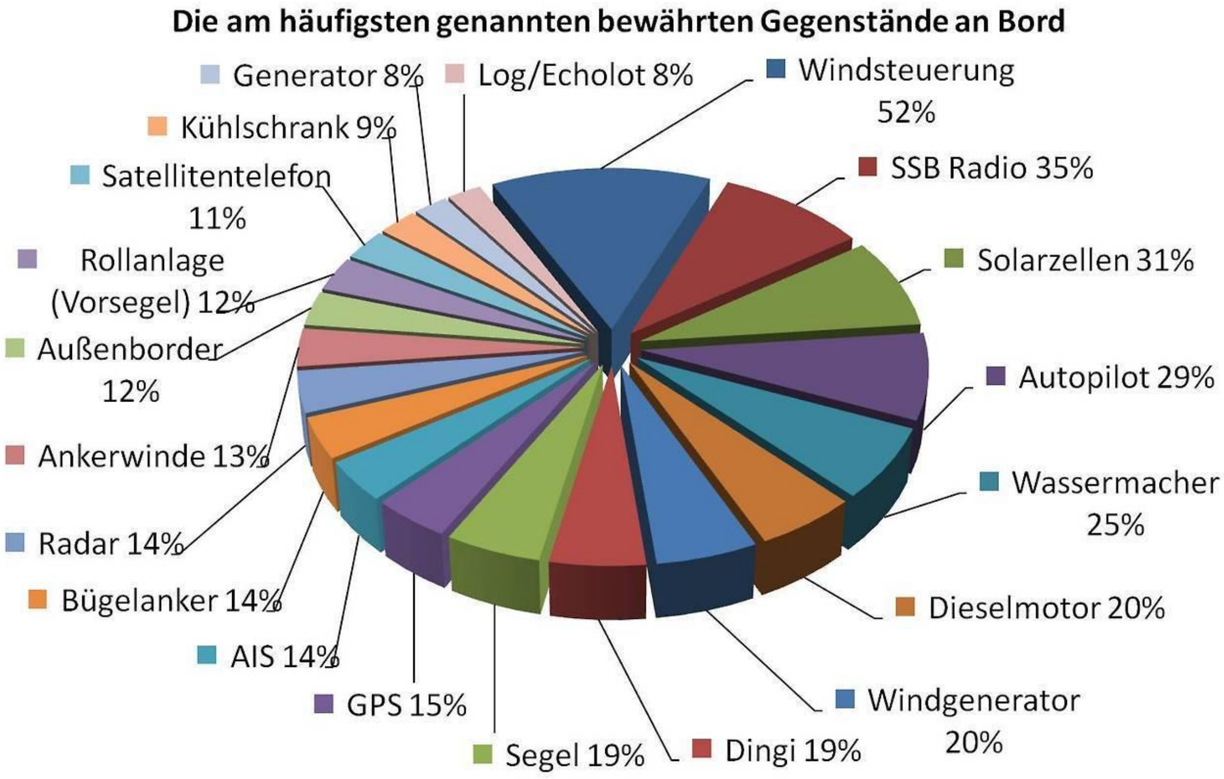
http://perceptualedge.com/files/GraphDesignIQ.html

text legibility



Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.

No Unjustified 3D: Examples

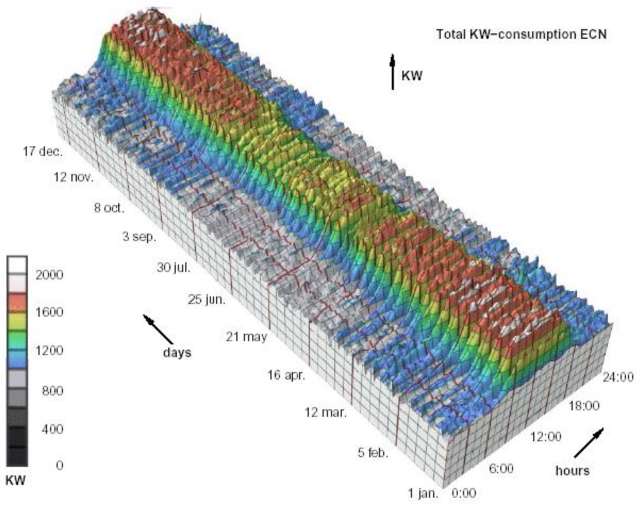


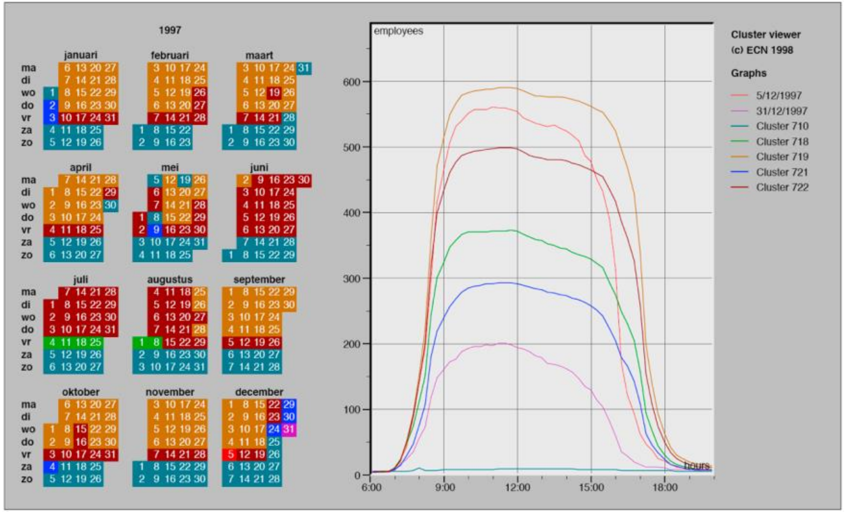
http://www.viz.wtf

No Unjustified 3D: Examples

extruded curves: detailed comparisons impossible

multiple views: calendar, superimposed 2D curves





Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis, 1999.

Justified 3D: Shape Perception

benefits outweigh costs when task is shape perception for 3D spatial data

– interactive navigation supports synthesis across many viewpoints

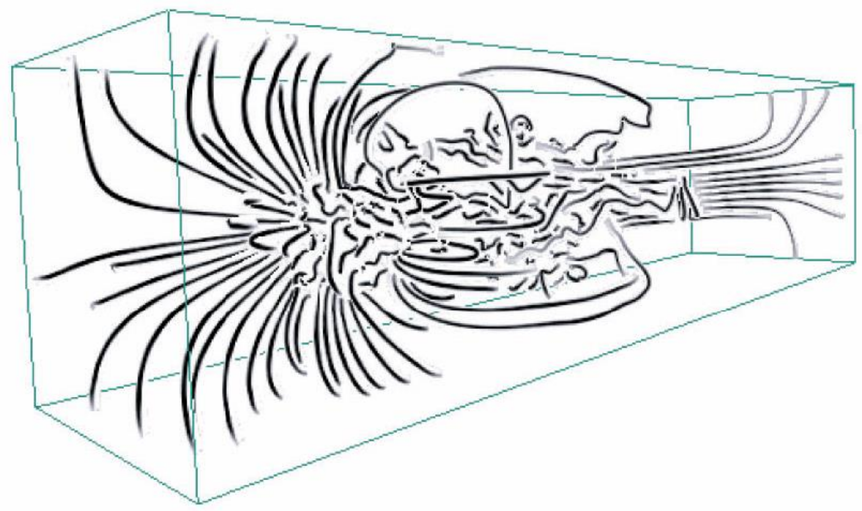
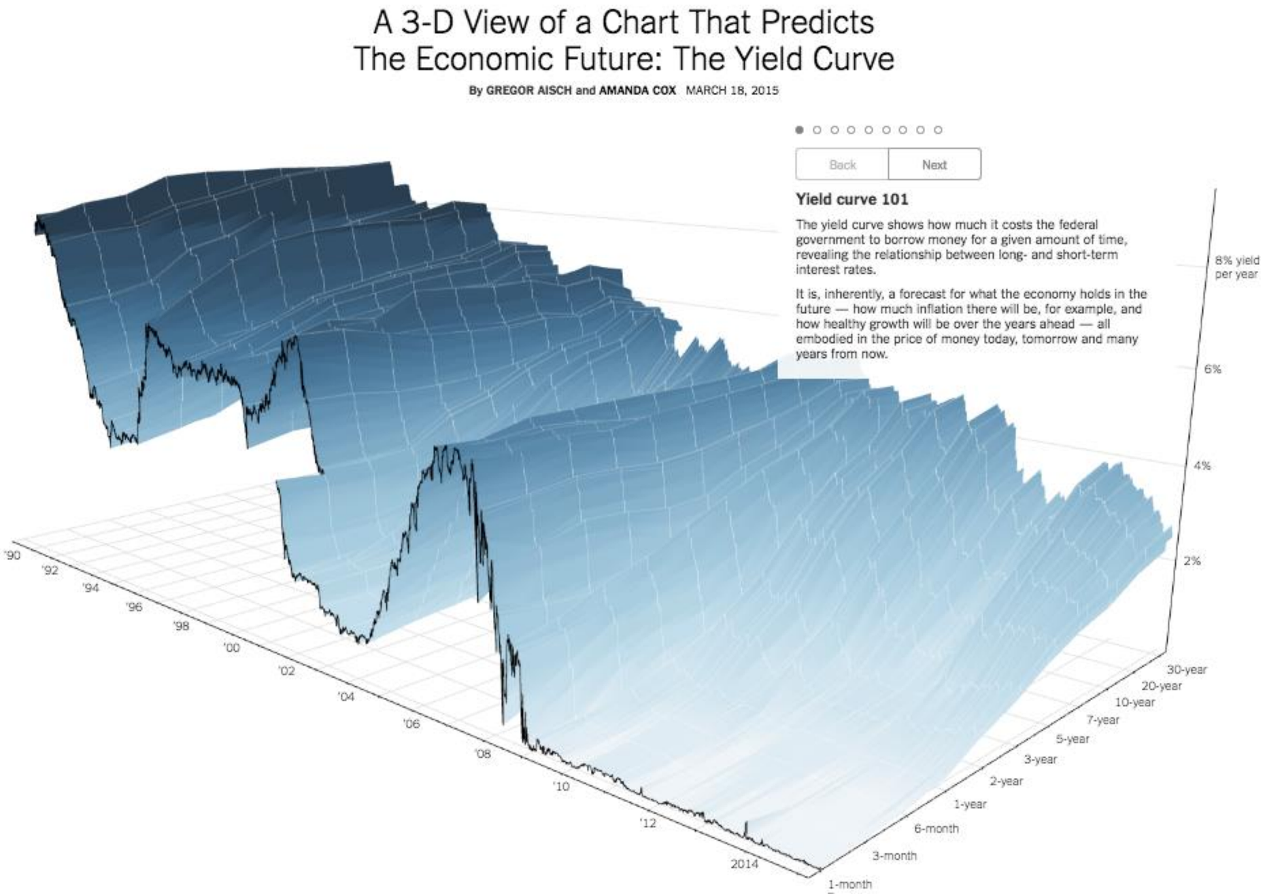


Image-Based Streamline Generation and Rendering. Li and Shen. IEEE TVCG, 2007.

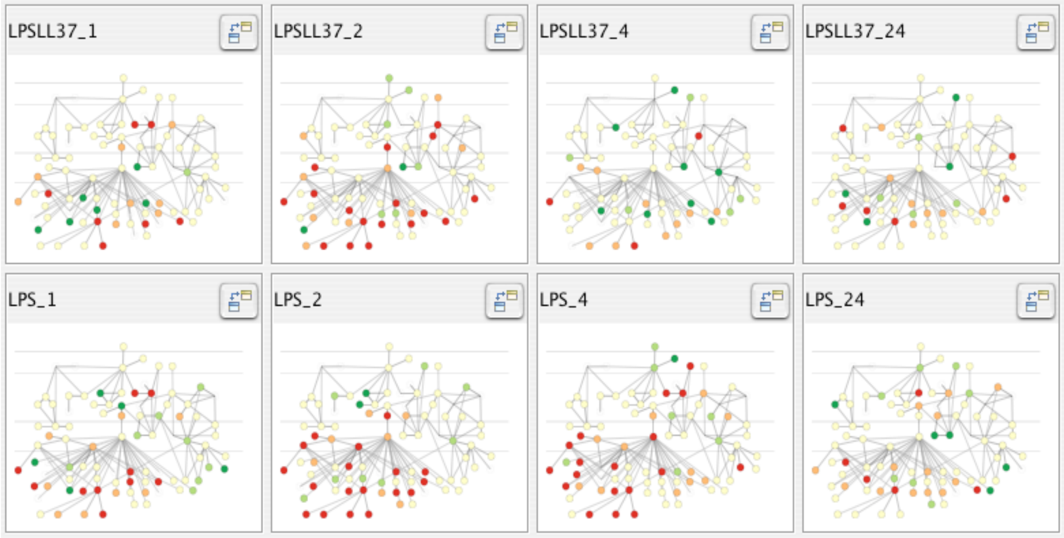
Justified 3D: Economic Growth Curve



Eyes Beat Memory

small multiples: one graph instance per experimental condition can be better then animation

– same spatial layout vs. color differently, by condition



Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG, 2008.

No Unjustified 2D

consider whether network data requires 2D spatial layout

- especially if reading text is central to task!
- arranging as network means lower information density and harder label lookup compared to text lists

benefits outweigh costs when topological structure/context important for task

- be especially careful for search results, document collections, ontologies

Final Words

overview first, zoom and filter, details on demand

– mantra from Shneiderman

start with focus on functionality

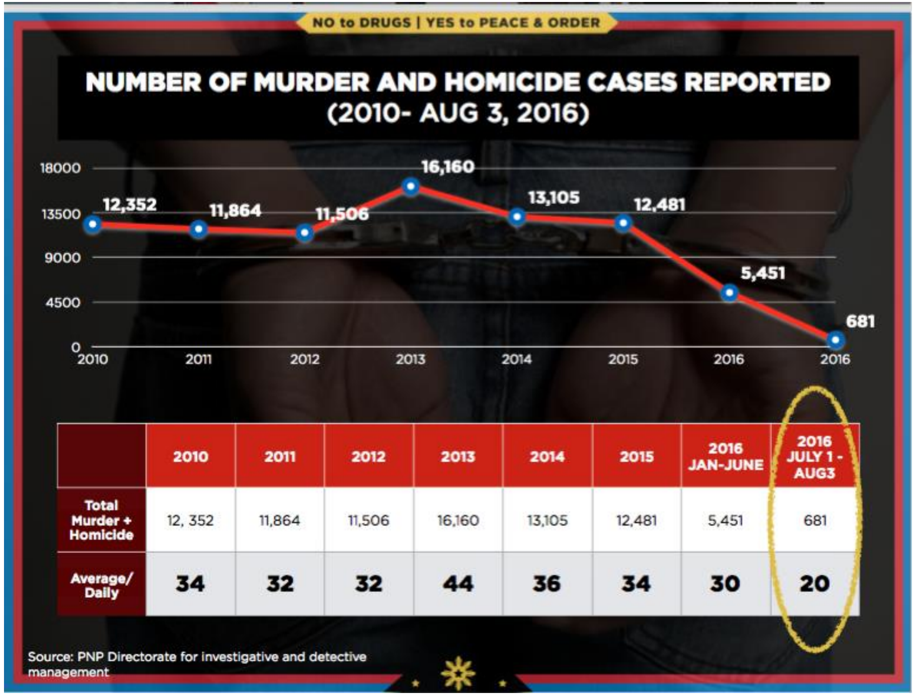
– straightforward to improve aesthetics later on, as refinement

dangerous to start with aesthetics

– usually impossible to add function retroactively

Beware!

Visualization can be exploited to confuse or to misinform!



	2010	2011	2012	2013	2014	2015	2016 JAN-JUNE	2016 JULY 1 - AUG3
Total Murder + Homicide	12, 352	11,864	11,506	16,160	13,105	12,481	5,451	681
Average/ Daily	34	32	32	44	36	34	30	20

http://www.viz.wtf

Acknowledgement

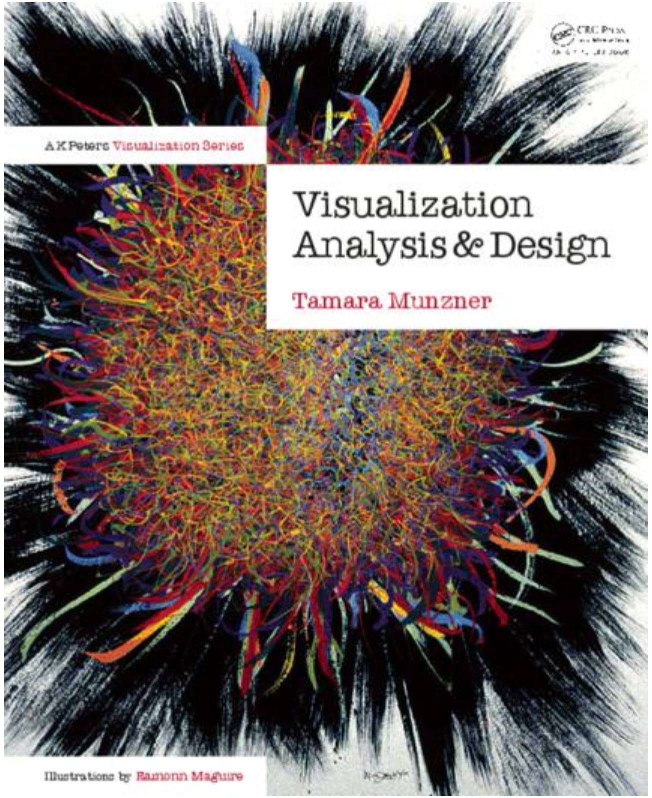
the source for this talk

- http://www.cs.ubc.ca/~tmm/talks.html

book page (including tutorial lecture slides)

- http://www.cs.ubc.ca/~tmm/vadbook
- http://www.crcpress.com/product/isbn/9781466508910
- 20% promo code for book+ebook combo: HVN17

illustrations: Eamonn Maguirethe



Visualization Analysis and Design.
Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.