POLITECNICO DI MILANO

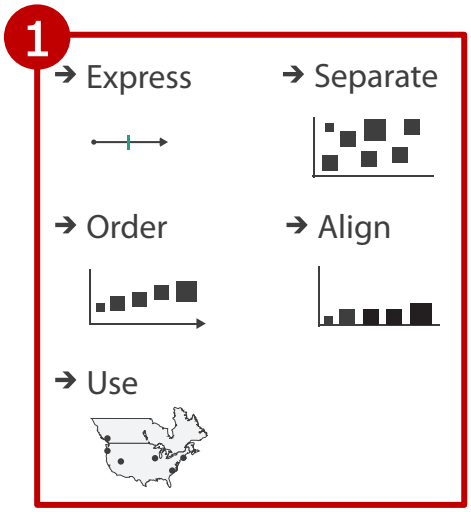
# Unit 3

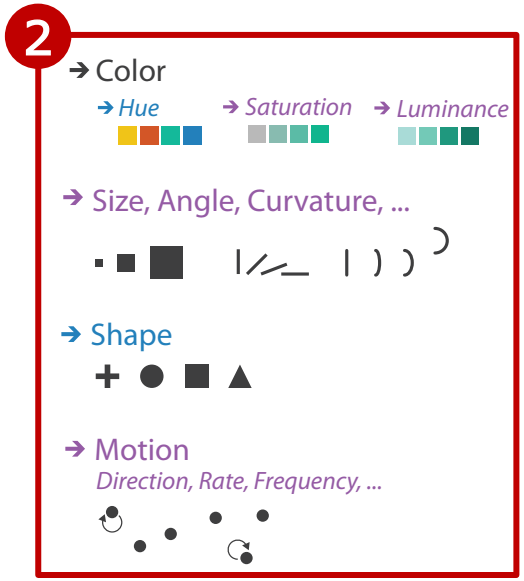
## Data and Results Visualization

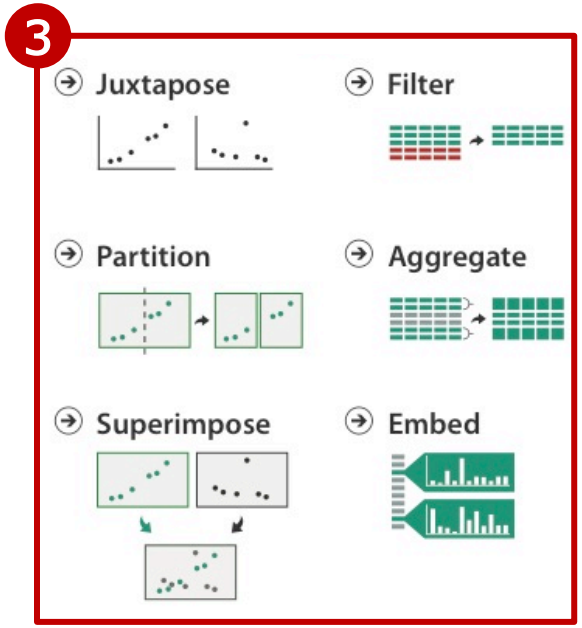
Daniele Loiacono

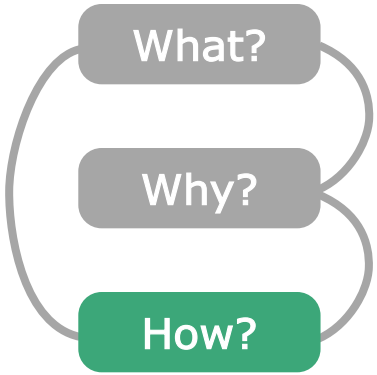
# GENERAL GUIDELINES

# Idiom design: an overview



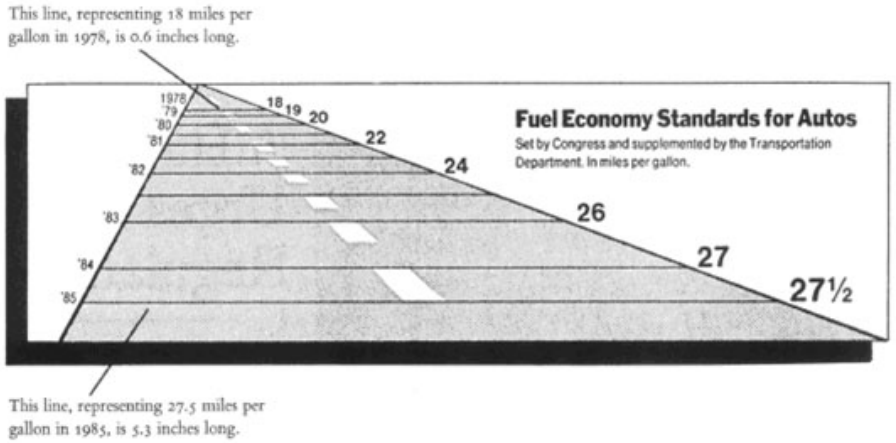




- Encode data
  - Arrange data in space 1
  - Map data to other channels 2
- Deal with complexity 3

# Tufte's Principles

- Graphical integrity (e.g., no data distortion)

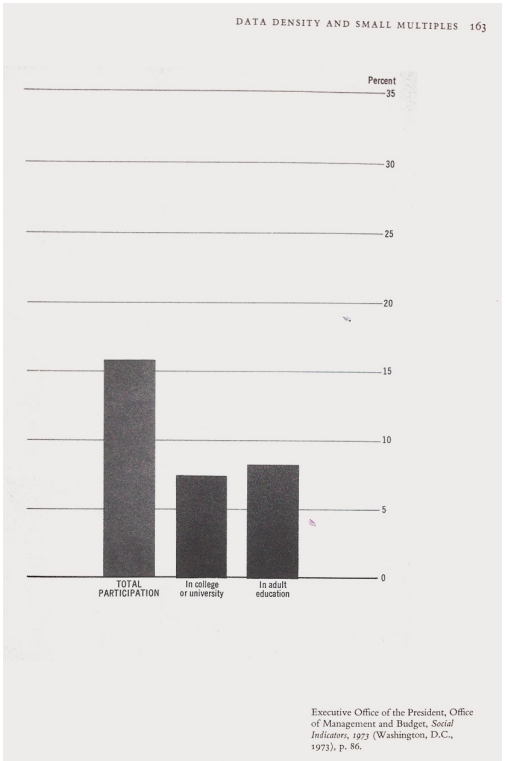


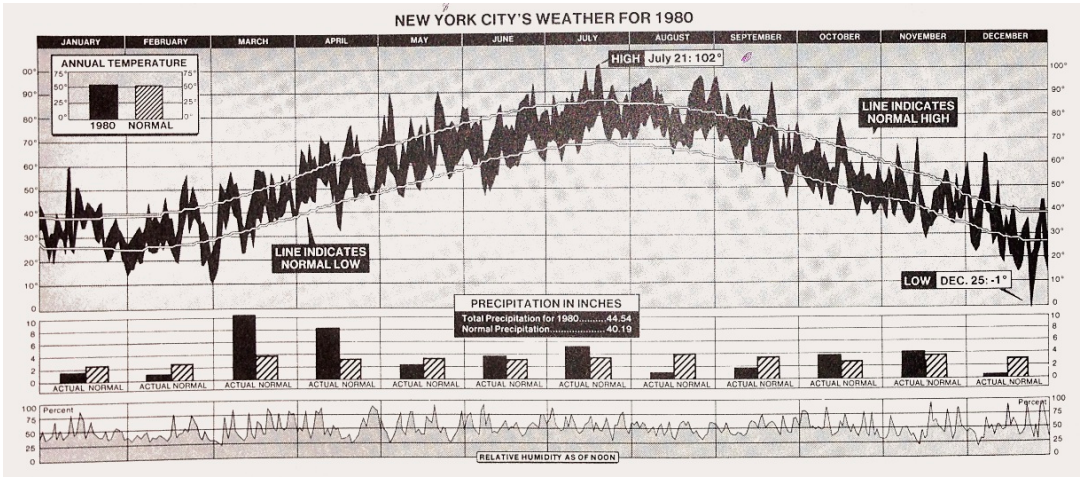
# Tufte's Principles

- Graphical integrity (e.g., no data distortion)
- Maximize data-ink

# Tufte's Principles

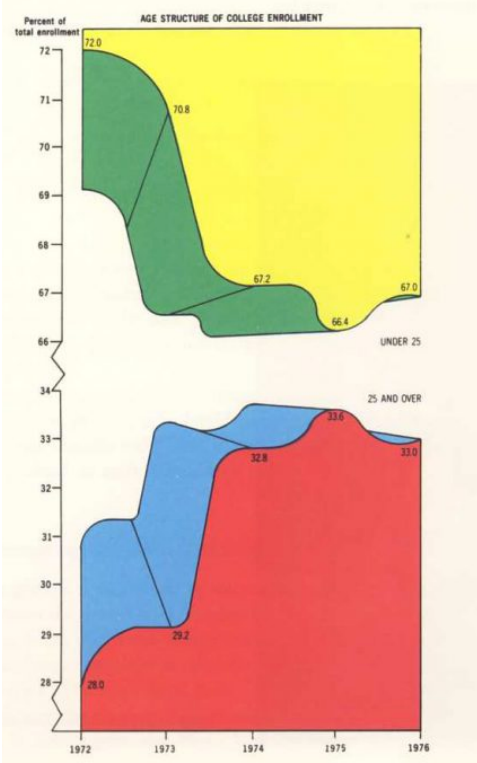
- Graphical integrity (e.g., no data distortion)
- Maximize data-ink
- Create high-density graphs



# Tufte's Principles

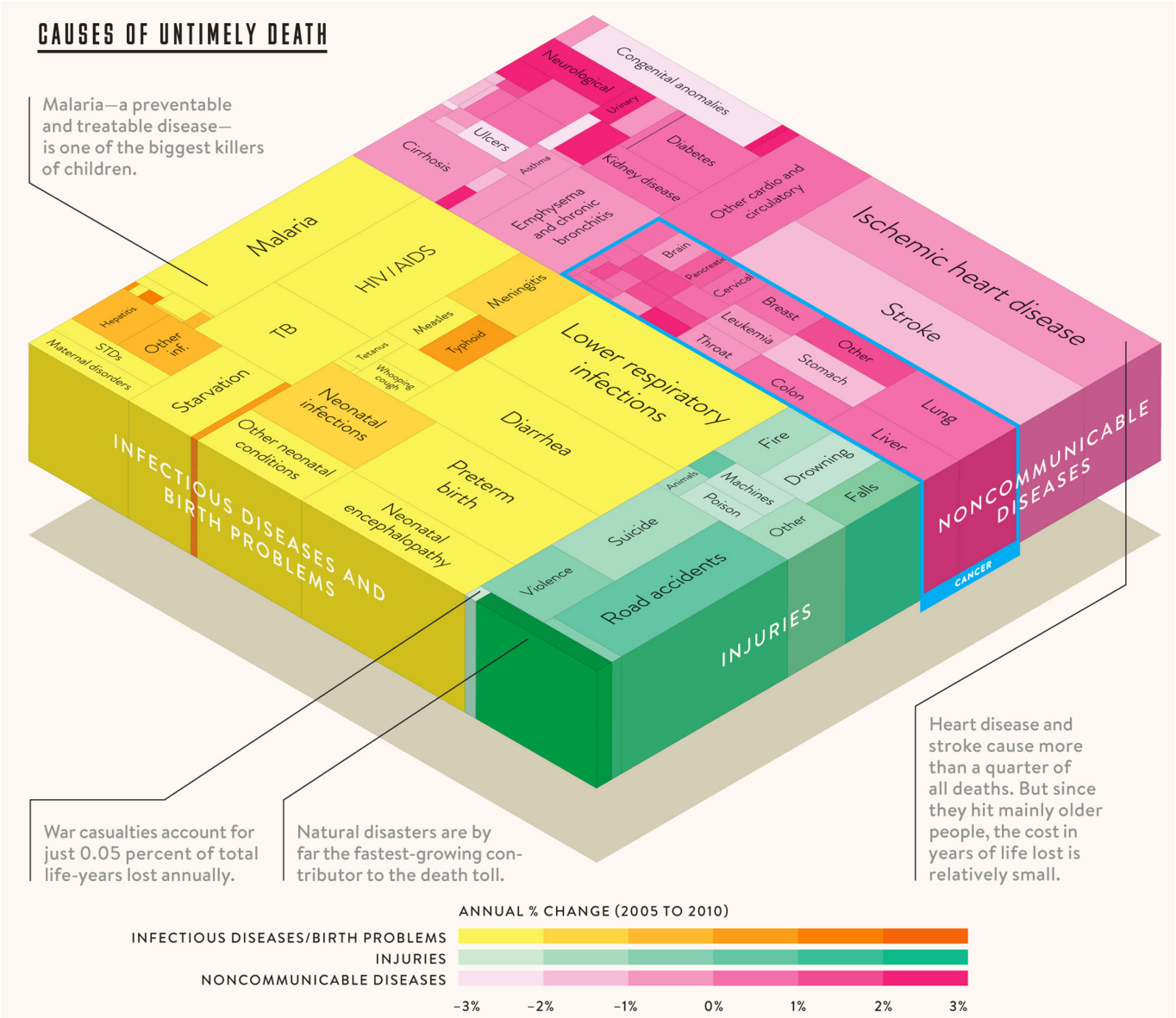
- Graphical integrity (e.g., no data distortion)
- Maximize data-ink
- Create high-density graphs
- Avoid unnecessary graphical effects



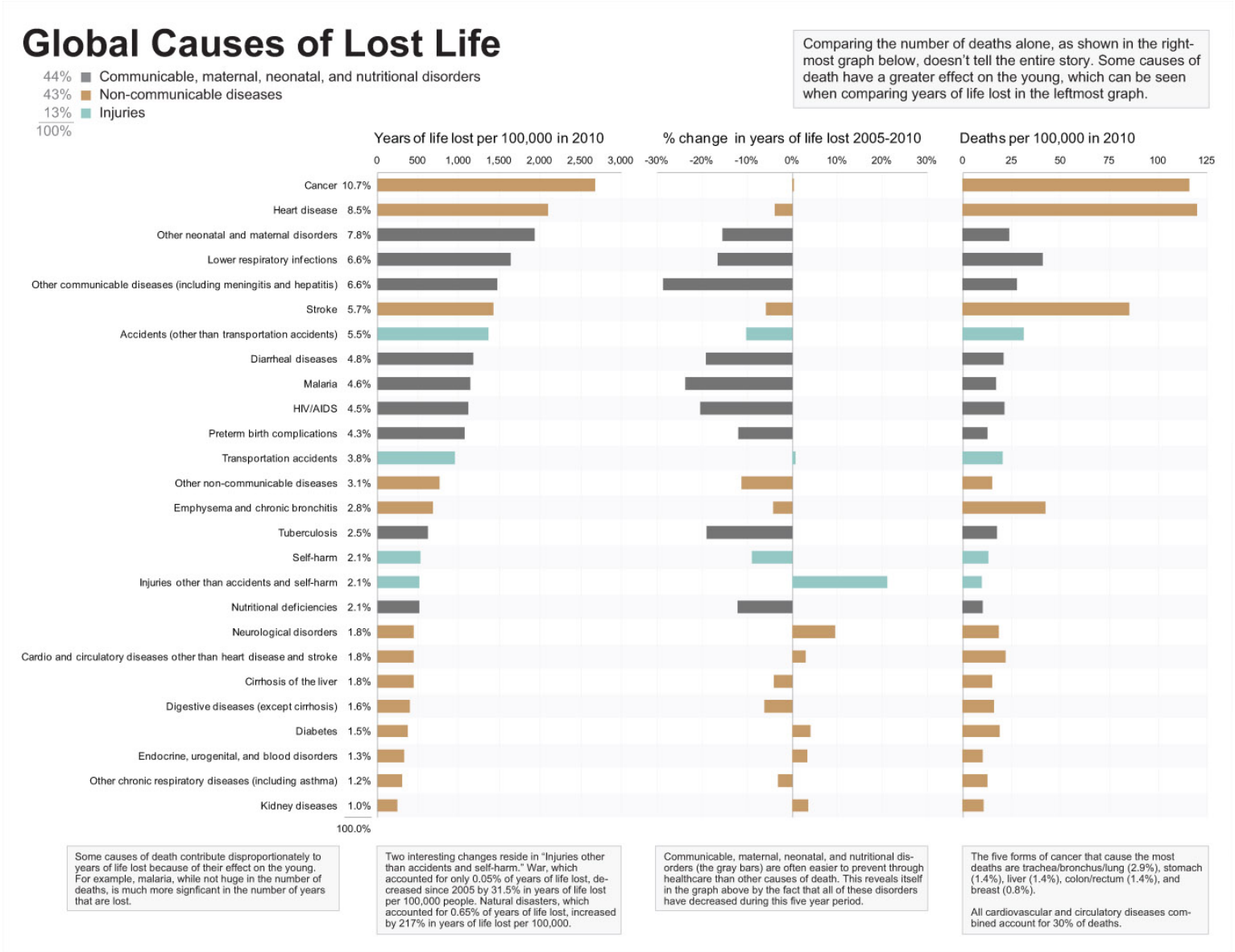
# No unjustified 3D

- Cons
  - Low accuracy on depth channel
  - Dangers of perspective distortion
  - Occlusion of information
  - Text legibility
- Pro
  - Better understanding of shapes

# No unjustified 3D: an example

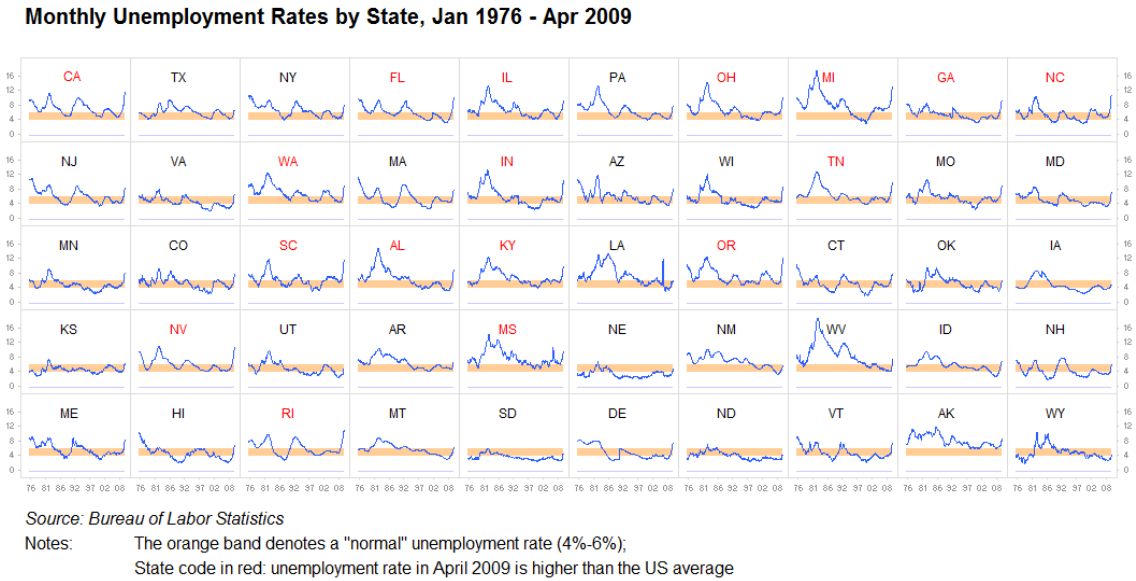


# No unjustified 3D: an example



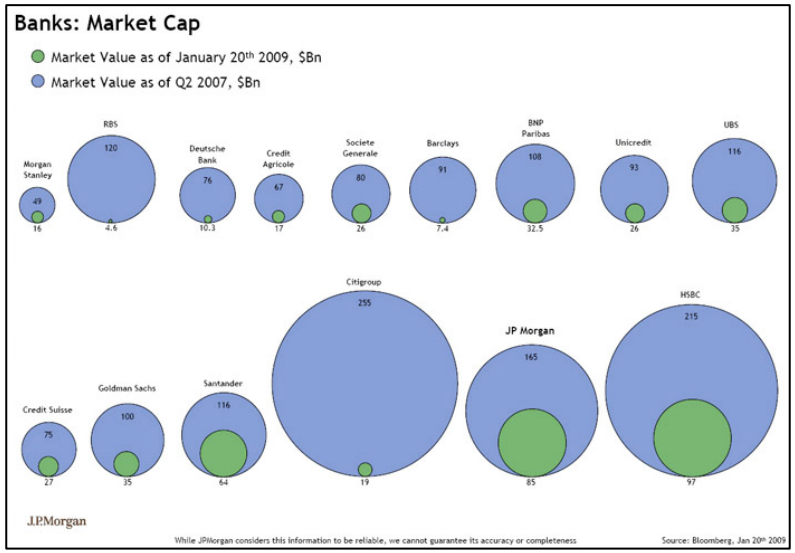
# Eyes beat memory

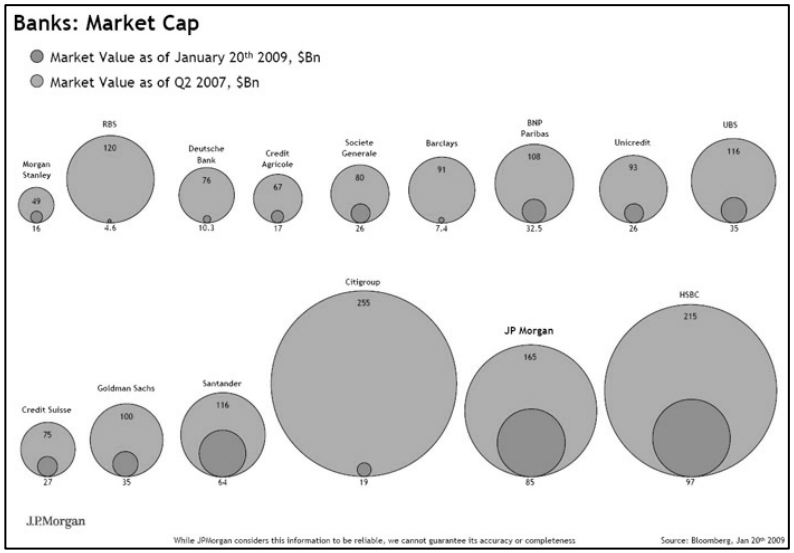
- When dealing with complex data do not **underestimate** visual perception and/or **overestimate** working memory.
- Transitions or not connected views lead to a bigger cognitive load
- Trade space for memory! (high-density graph)
  - Multiple connected views
  - Superimpose
  - Aggregation



# Get it right in black and white

- Make sure that most important data is visible also when image is transformed to black and white (**check it!**).
  - use space and luminance/saturation as major source of information
  - others channels are to be used for less important data

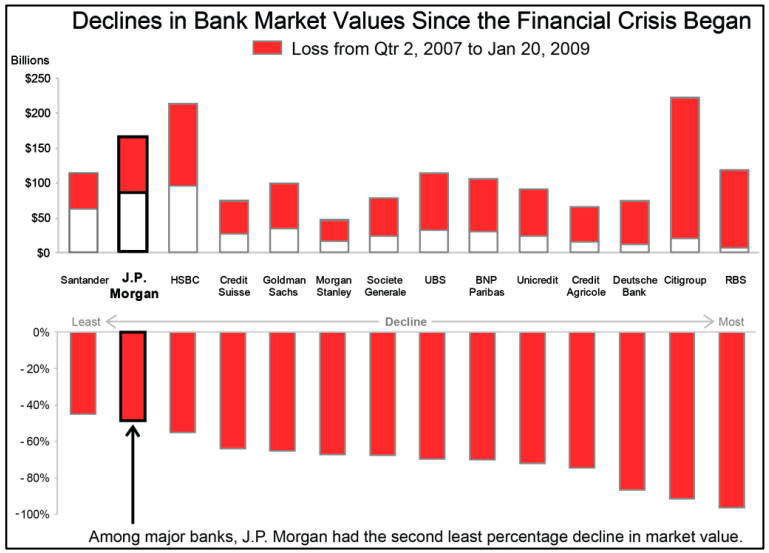


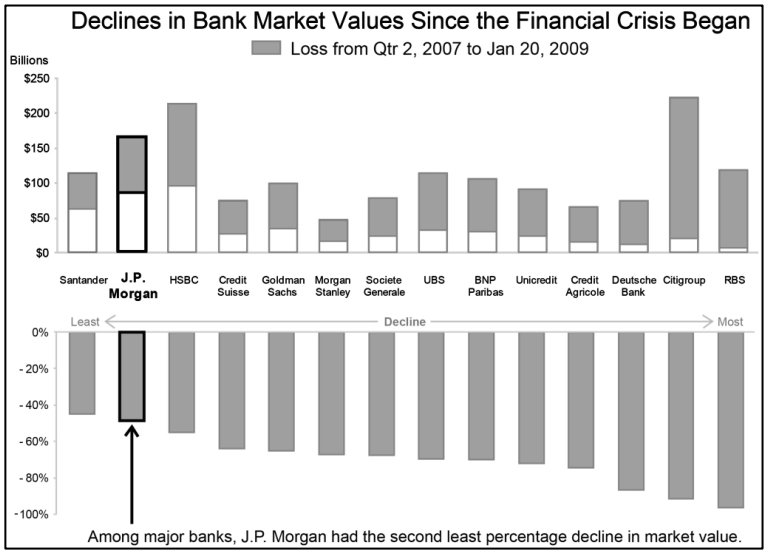




Stephen Few: http://www.perceptualedge.com/example18.php

# Get it right in black and white

- Make sure that most important data is visible also when image is transformed to black and white (**check it!**).
  - use space and luminance/saturation as major source of information
  - others channels are to be used for less important data

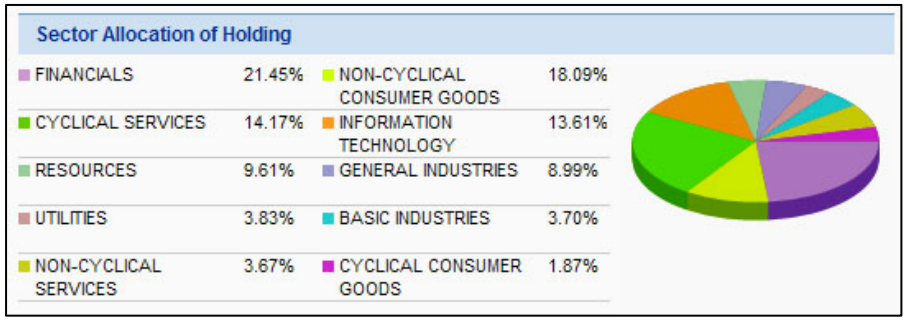


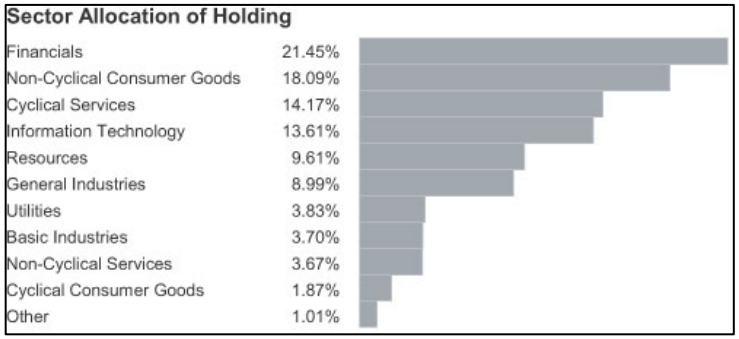




Stephen Few: http://www.perceptualedge.com/example18.php

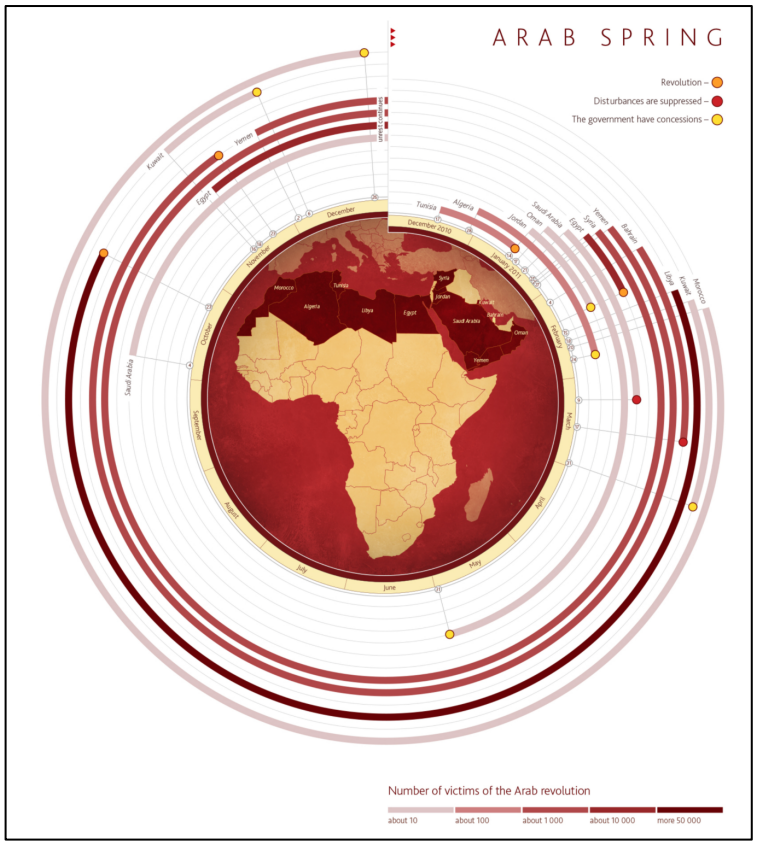
# Function first!

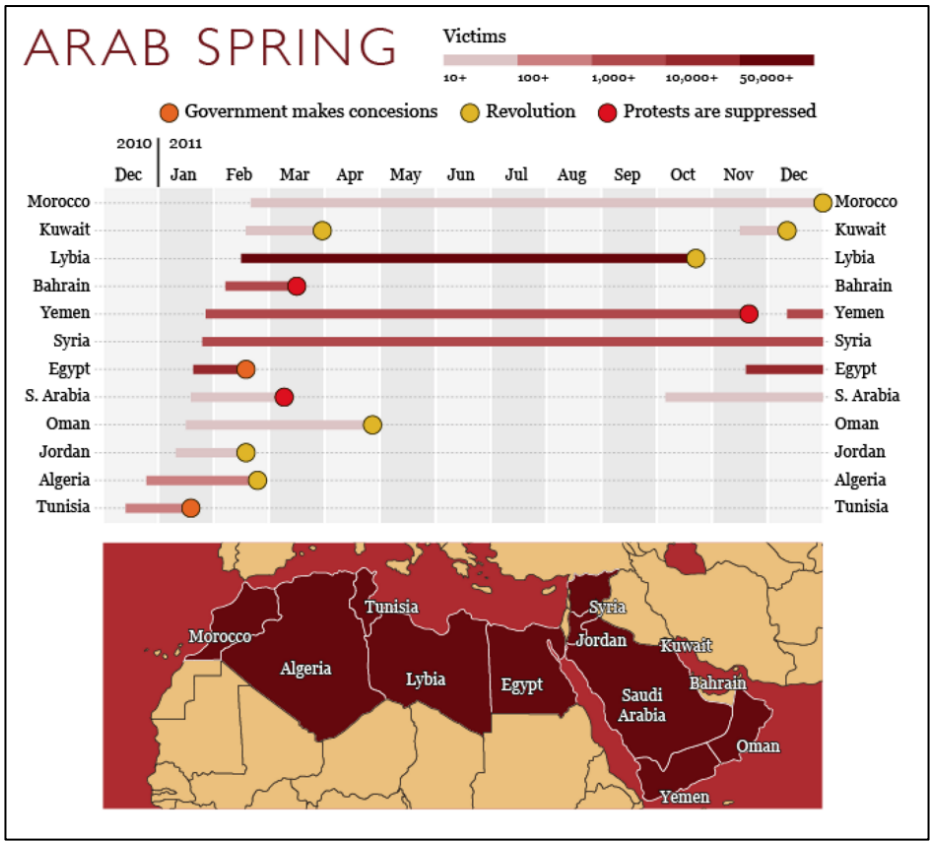
- A beautiful visualization design leads to more engagement.
- However, function is **essential** and form can be more easily **improved iteratively.**



# Function vs form: an example




http://www.thefunctionalart.com/2015/02/redesigning-circular-timeline.html

# VISUALIZATION OF TABLES

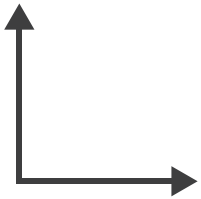
# How to use space?

# Space and quantitative attributes

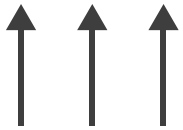
- Express values with position

- How to design the axis?

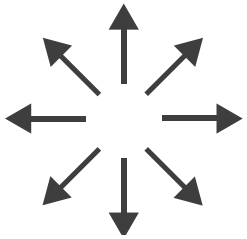
→ Rectilinear



→ Parallel



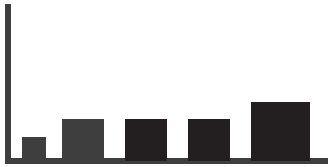
→ Radial



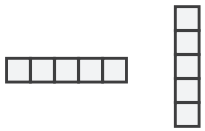
# Space and categorical attributes

- Map categorical key(s) to separated regions

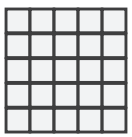
- Choose a proper alignment



→ *1 Key*
*List*



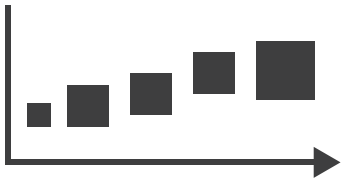
→ *2 Keys*
*Matrix*



→ *3 Keys*
*Volume*

→ *Many Keys*
*Recursive*

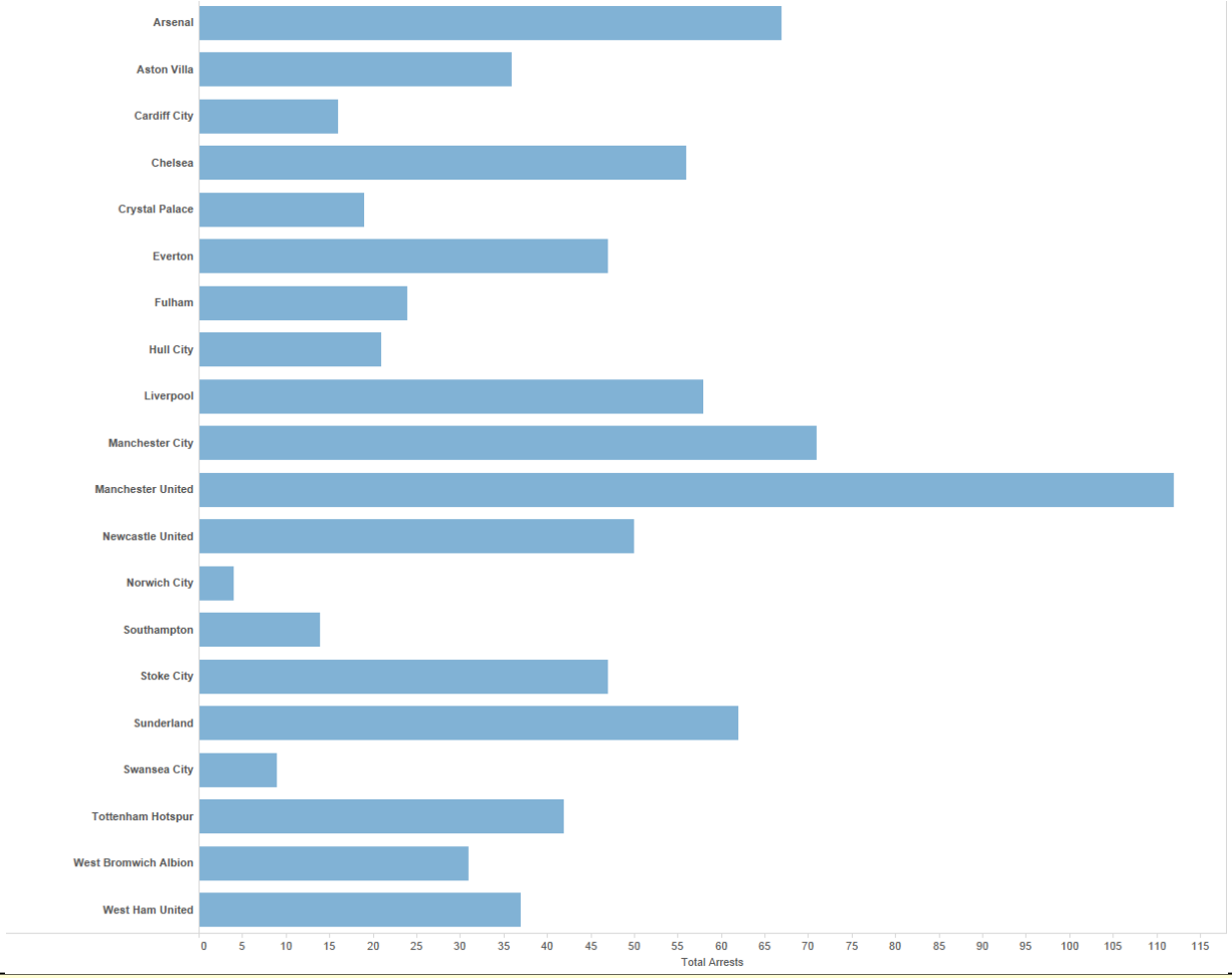
- Choose a proper ordering



# Space and categorical attributes: example


Separated but not aligned/ordered: difficult to compare

# Space and categorical attributes: example

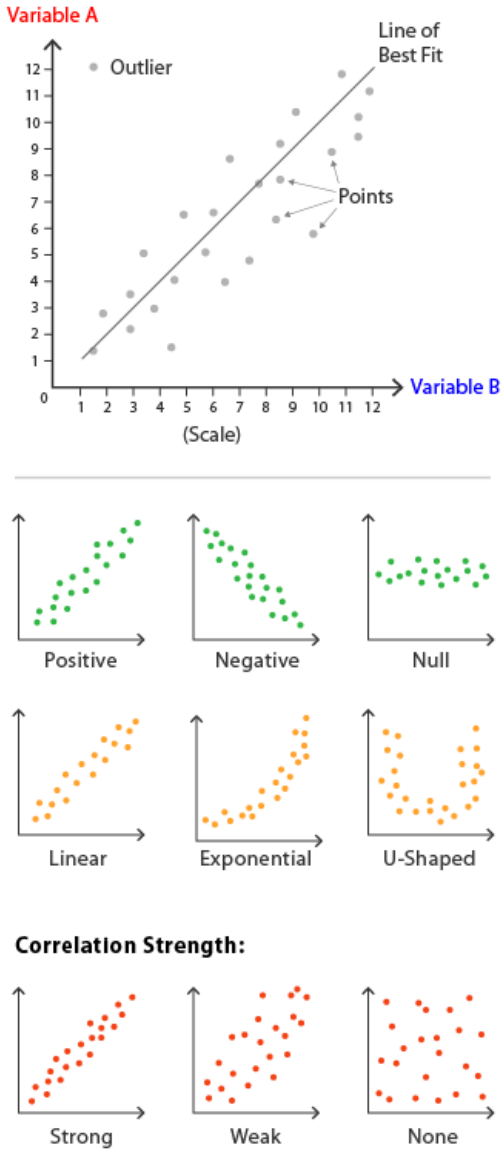



Separated and aligned but not ordered: which is the 4th largest?

# Idioms for quantitative attributes

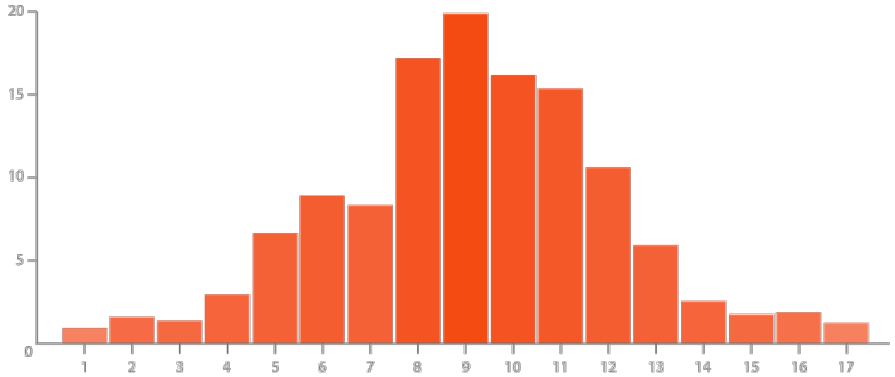
# Scatter plot: anatomy

- What?
  - 2 quantitative attribute
- Why?
  - Correlation and distribution
  - Identify outliers, patterns, and clusters
- Remarks
  - Scale up to hundreds of items
  - Color and size can be used to encode two additional categorical attributes (bubble plot)



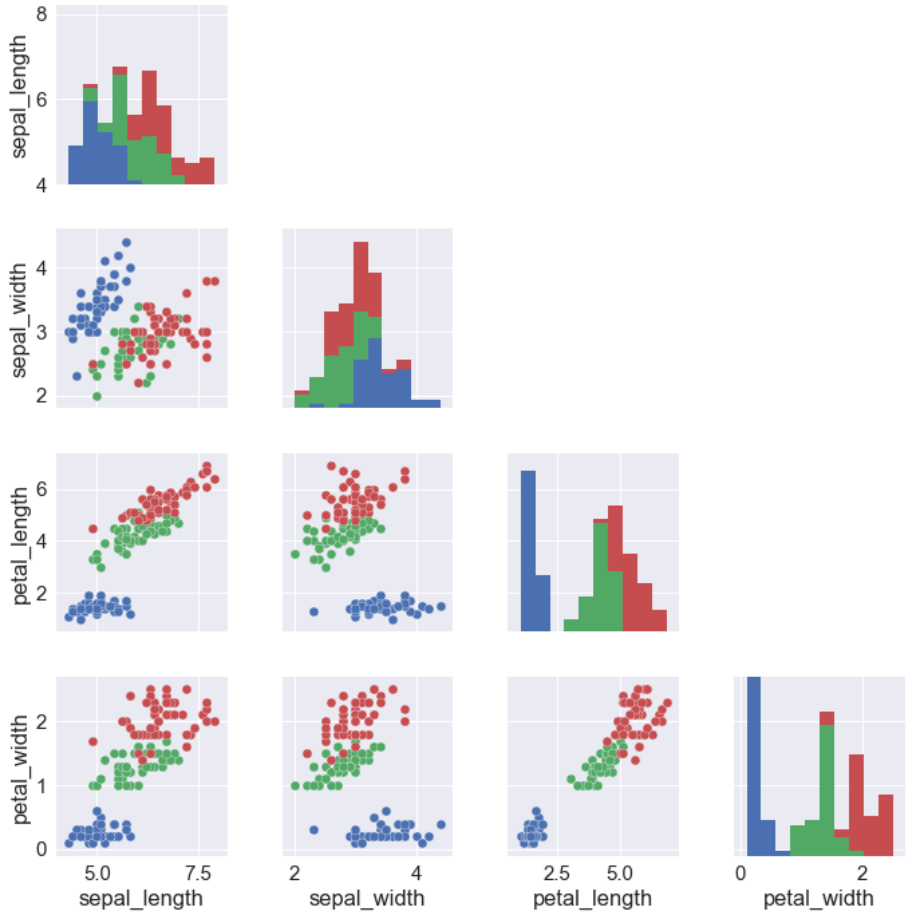
# Histogram: anatomy

- What?
  - 1 quantitative attribute
- Why?
  - Identify distribution, patterns, and range
- Remarks
  - A line (or area) can plotted to show computed density function
  - Items can be plotted with dots



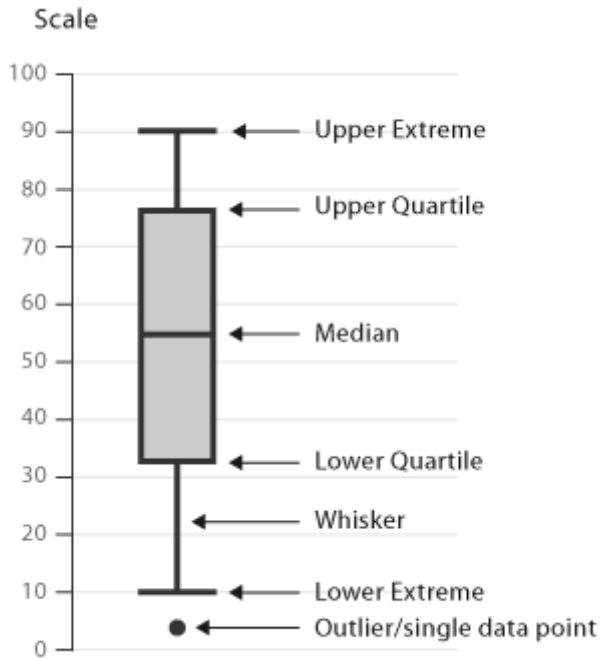
# Scatter plot matrix: anatomy

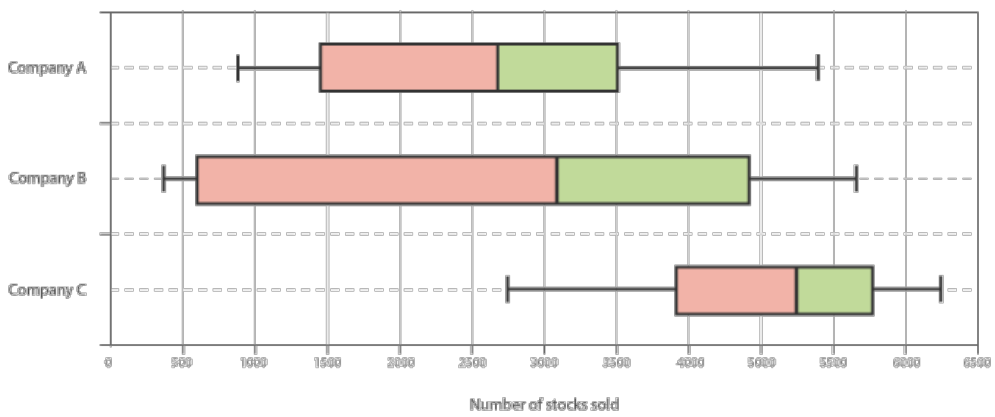
- What?
  - N quantitative attribute
- Why?
  - Correlation and distribution
  - Identify outliers, patterns, and clusters
- Remarks
  - Scale up one dozen of attributes and hundreds of items
  - Possible to remove upper triangle (or map to different plot)



# Box plot: anatomy

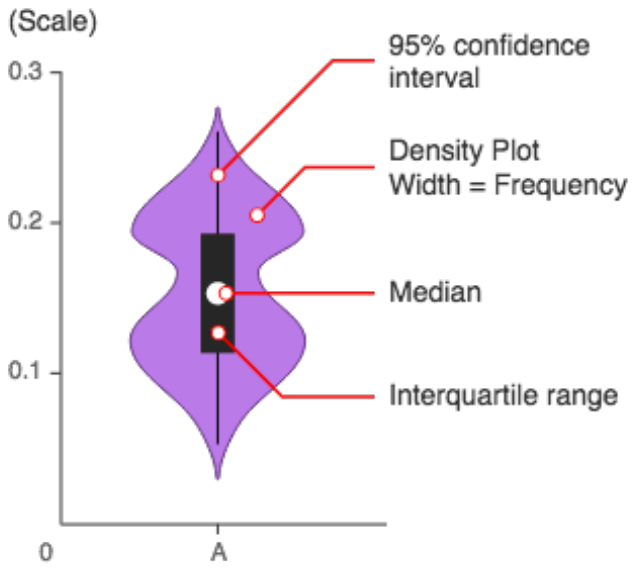
- What?
  - N quantitative attributes
  - (or 1 quantitative and 1 categorical key)
- Why?
  - Distribution
  - Identify outliers, extremes, etc.
- Remarks
  - Color can encode categorical attribute
  - Grouping possible
  - Alternative glyphs

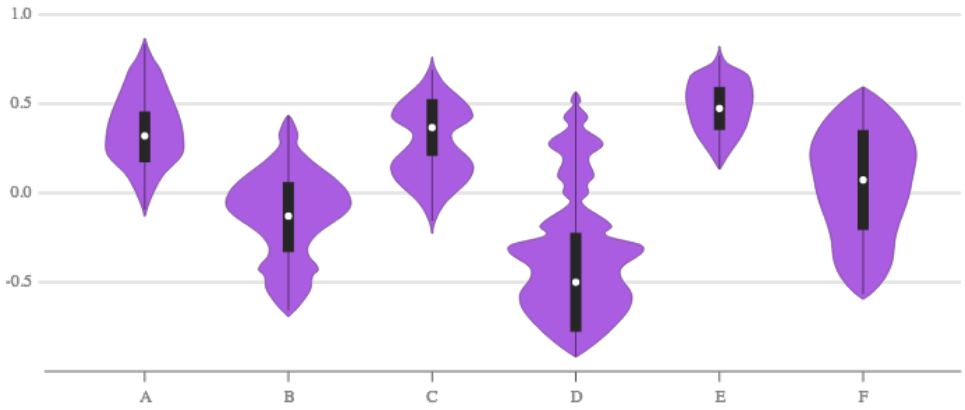


# Violin plot: anatomy

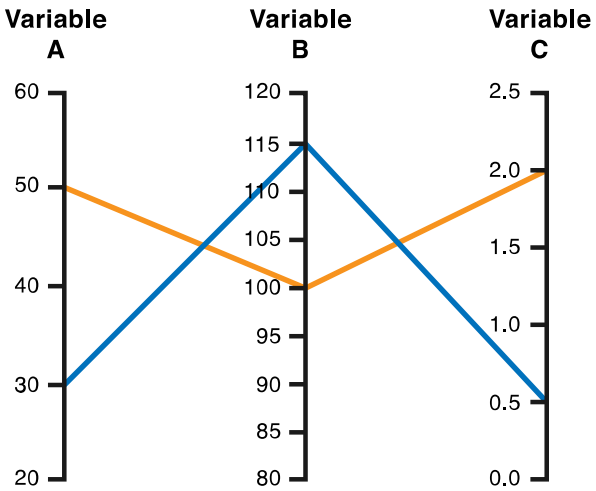
- What?
  - N quantitative attributes
  - (or 1 quantitative and 1 categorical key)
- Why?
  - Distribution
  - Identify range
- Remarks
  - Color can encode categorical attribute (if binary split is possible)
  - Similar to letter-value plot



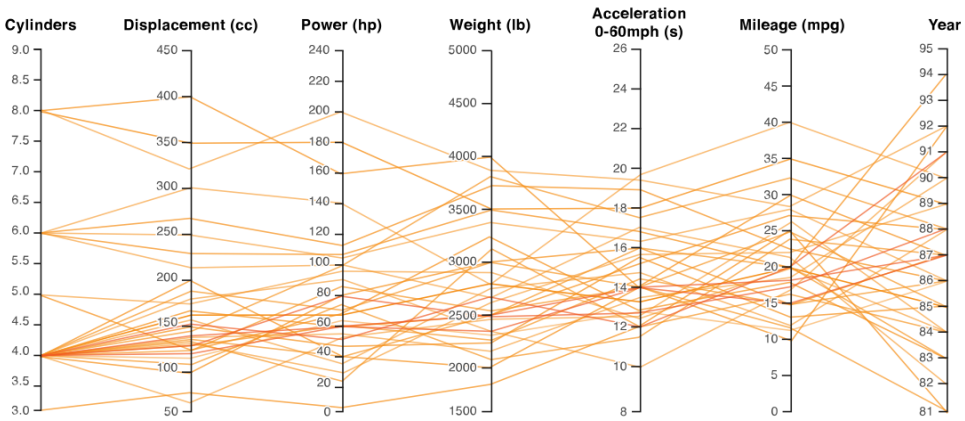



# Parallel coordinates: anatomy

- What?
  - N quantitative/ordered attributes
- Why?
  - Correlation
  - Identify outliers, range and patterns
- Remarks
  - Scale up to hundreds of items and a dozen of attributes
  - Color(hue) can encode additional categorical attribute

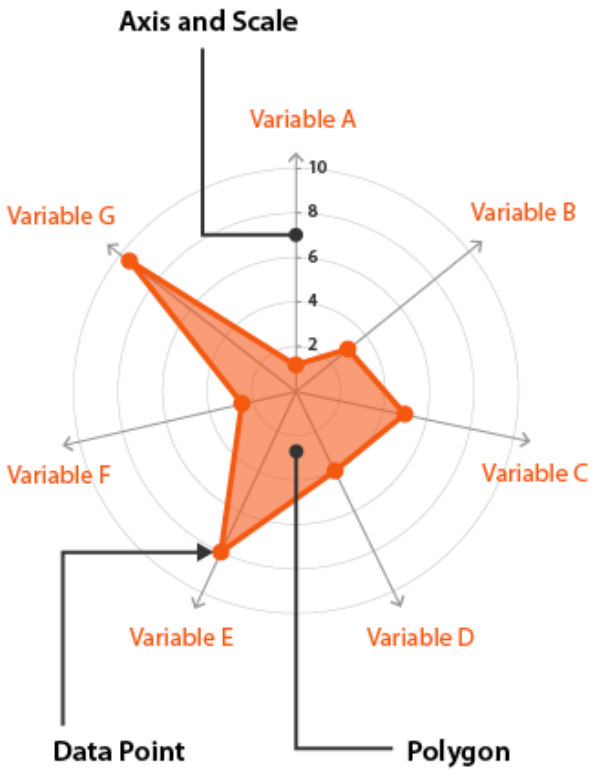



| Data | | | |
|---|---|---|---|
| | Variable A | Variable B | Variable C |
| Item 1 | 50 | 100 | 2.0 |
| Item 2 | 30 | 115 | 0.5 |



# Radar chart: anatomy

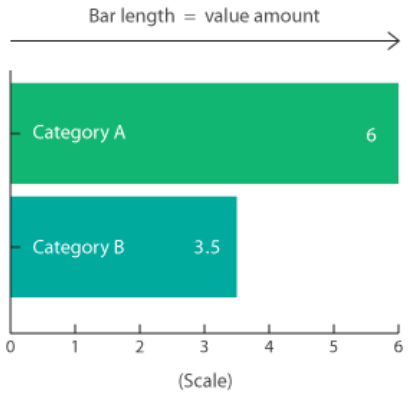
- What?
  - N quantitative attribute
- Why?
  - Identify patterns
  - Compares values
- Remarks
  - Scale up to a dozen attributes
  - Color (hue) can encode an additional categorical key (up to 3-4 values)
  - Not filled version to improve readability



# Idioms for one categorical key

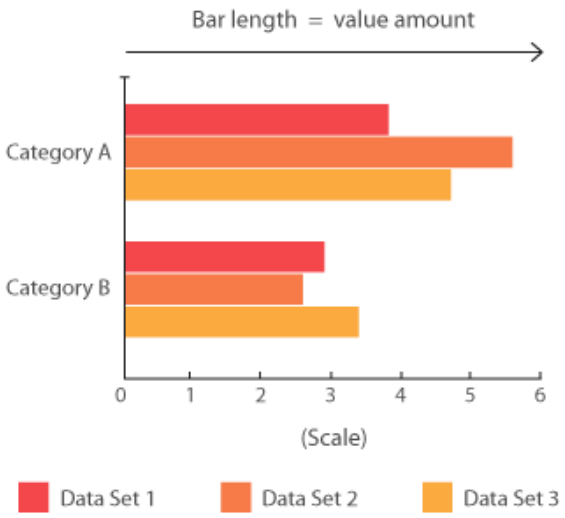
# Bar: anatomy

- What?
  - 1 quantitative attribute
  - 1 categorical key
- Why?
  - Compare/Lookup values
  - Identify extremes
- Remarks
  - Scale up to ~100 bars
  - Key vs values ordering
  - Not suitable for trends



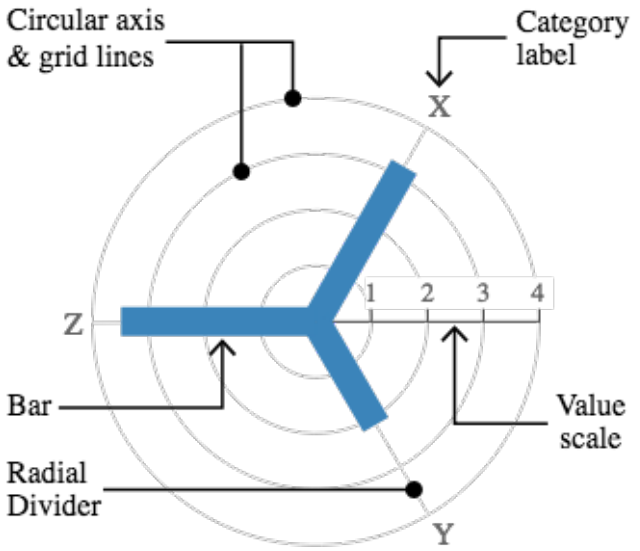
# Multi-set bar: anatomy

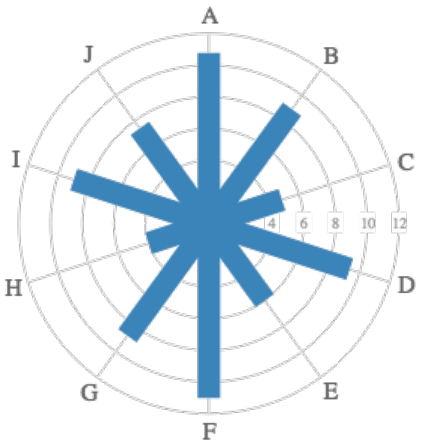
- What?
  - 1 quantitative attribute
  - 2 categorical keys
- Why?
  - Compare values
  - Identify patterns
- Remarks
  - Scale up to ~100 bars
  - The grouping key affects patterns/comparison focus

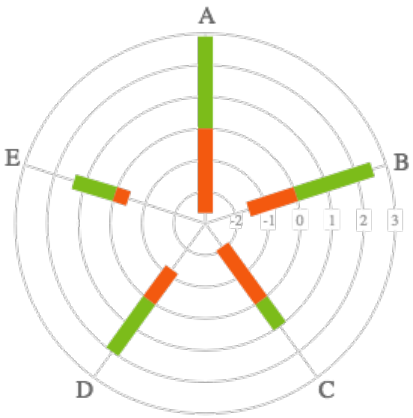


# Radial Column bar: anatomy

- What?
  - 1 quantitative attribute
  - 1 categorical key
- Why?
  - Compare/Lookup values
  - Identify pattern
- Remarks
  - Allow stacked design
  - Less dense than linear layout (scale up to ~20 bars)
  - Less accuracy to compare values

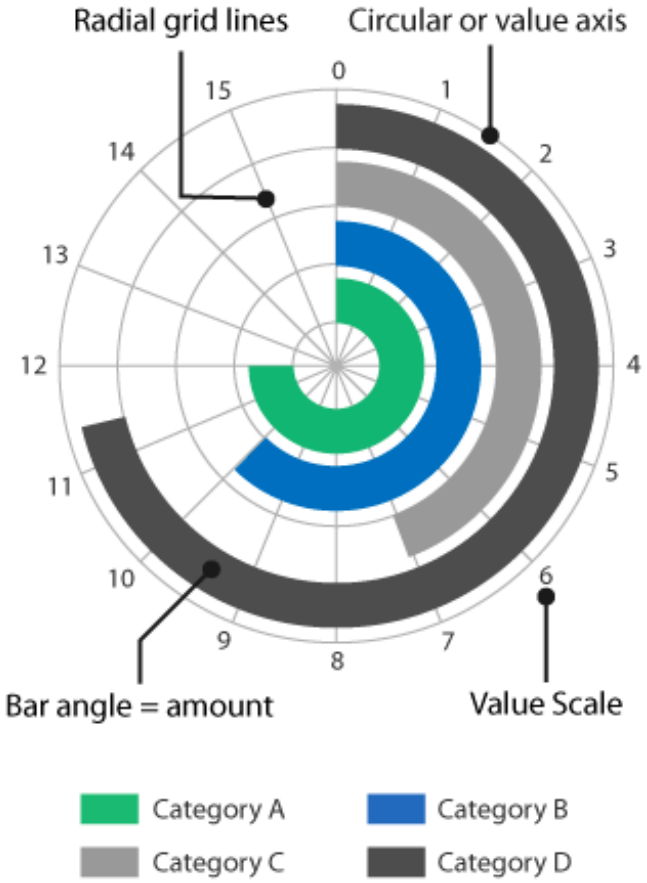


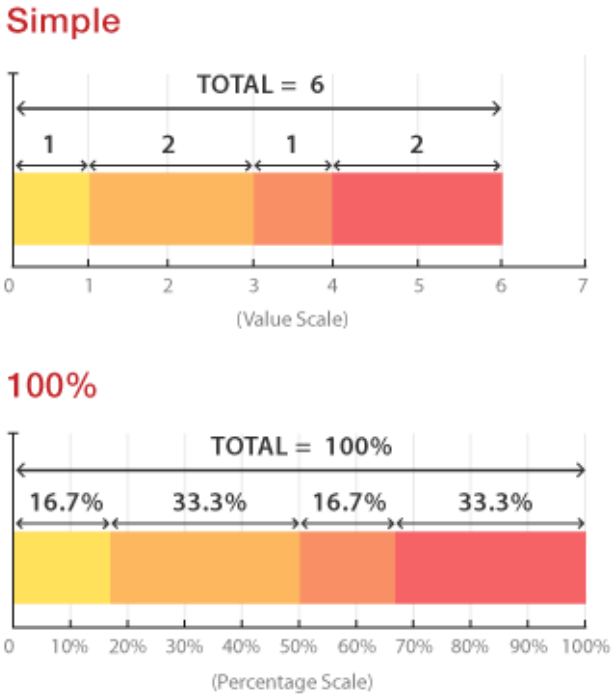
# Radial bar: anatomy

- What?
  - 1 quantitative attribute
  - 1 categorical key
- Why?
  - Compare/Lookup values
- Remarks
  - Much less dense than linear layout (max 10 bars)
  - Can be used for intervals (from a start value to an end value)



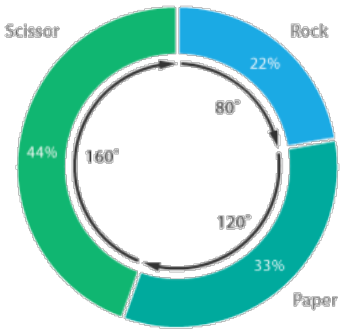
# Stacked bar: anatomy

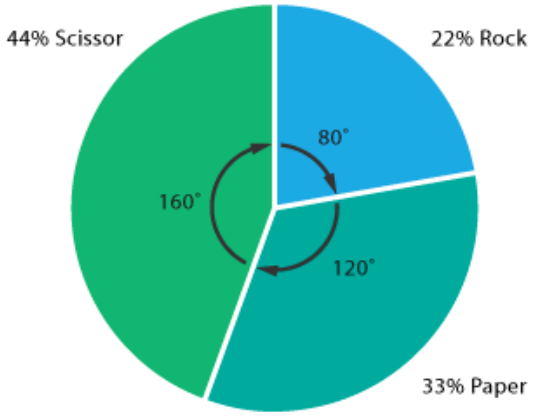
- What?
  - 2 categorical keys
  - 1 quantitative attribute
- Why?
  - Part-to-whole
  - Compare/lookup values
  - Identify patterns
- Remarks
  - Absolute vs Normalized
  - Ordering of secondary key affect accuracy



# Pie chart: anatomy

- What?
  - 1 quantitative attribute
  - 1 categorical key
- Why?
  - Part-to-whole
- Remarks
  - Less dense and accurate than linear layout
  - Central area can be removed (donut chart)

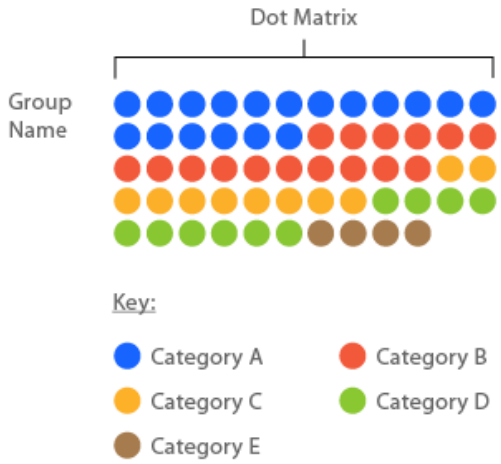



| Data | | | |
|---|---|---|---|
| Rock | Paper | Scissor | TOTAL |
| 2 | 3 | 4 | 9 |
| To calculate percentages | | | |
| 2/9=22% | 3/9=33% | 4/9=44% | 100% |
| Degrees for each "pie slice" | | | |
| (2/9) x 360 = 80° | (3/9) x 360 = 120° | (4/9) x 360 = 160° | 360° |

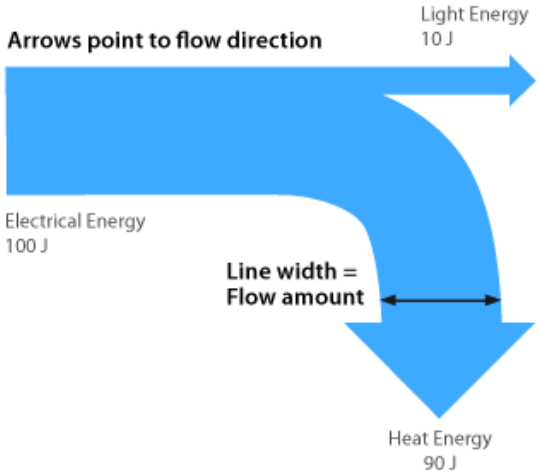
# Dot matrix chart: anatomy

- What?
  - 1 categorical attribute
- Why?
  - Part-to-whole
  - Identify patterns
- Remarks
  - Spatial arrangement often used to encode a categorical key



# Sankey diagram: anatomy

- What?
  - 1 quantitative attribute
  - 1 categorical key
- Why?
  - Part-to-whole
- Remarks
  - Color (hue) can encode an additional categorical attribute



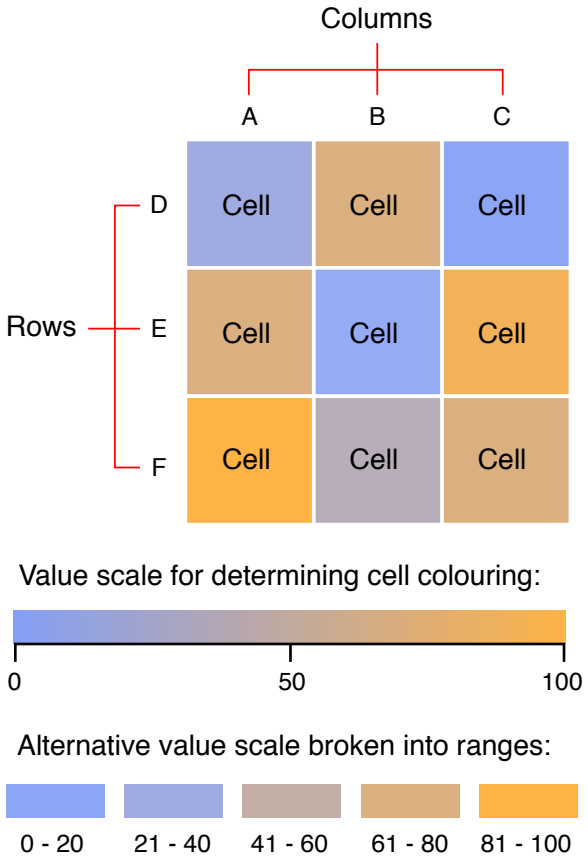
# Word cloud: anatomy

- What?
  - 1 categorical attribute (text)
  - 1 quantitative (frequency)
- Why?
  - Distribution of keyword
  - Summarization
- Remarks
  - Poor accuracy
  - Bias due to word length and structure

# Idioms for two categorical keys

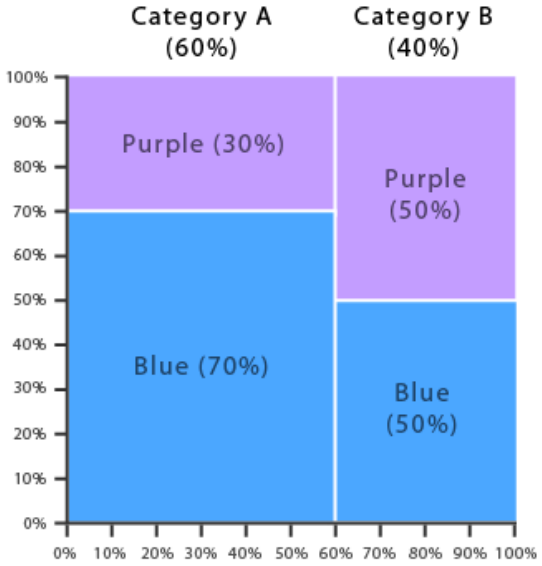
# Heatmap: anatomy

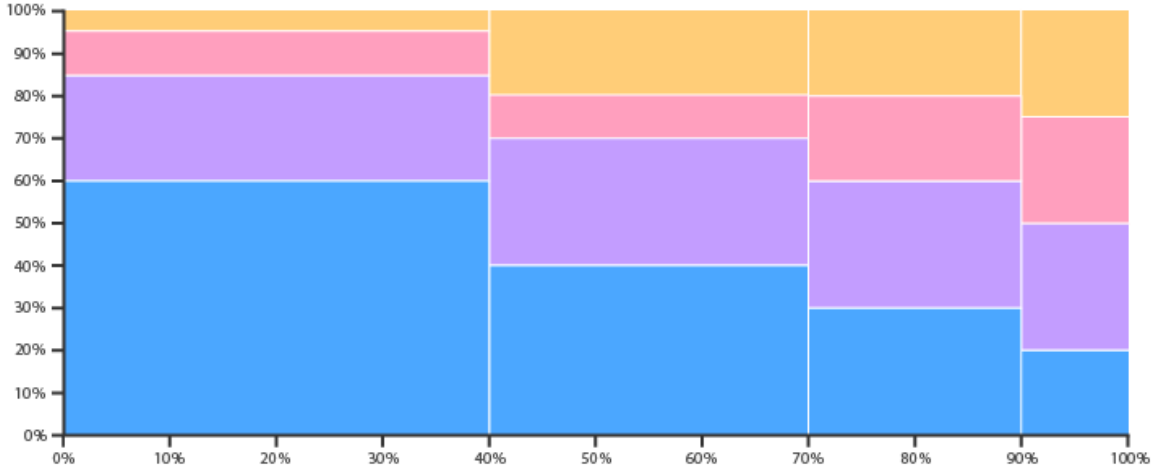
- What?
  - 2 categorical keys
  - 1 quantitative attribute
- Why?
  - Discover patterns, outliers
  - Correlation
- Remarks
  - Scale up to ~1M of items
  - Keys ordering affects the discoverability of patterns



# Marimekko chart: anatomy

- What?
  - 2 categorical keys
  - 1 quantitative attribute
- Why?
  - Part-to-whole
- Remarks
  - Value-based ordering of keys might work better



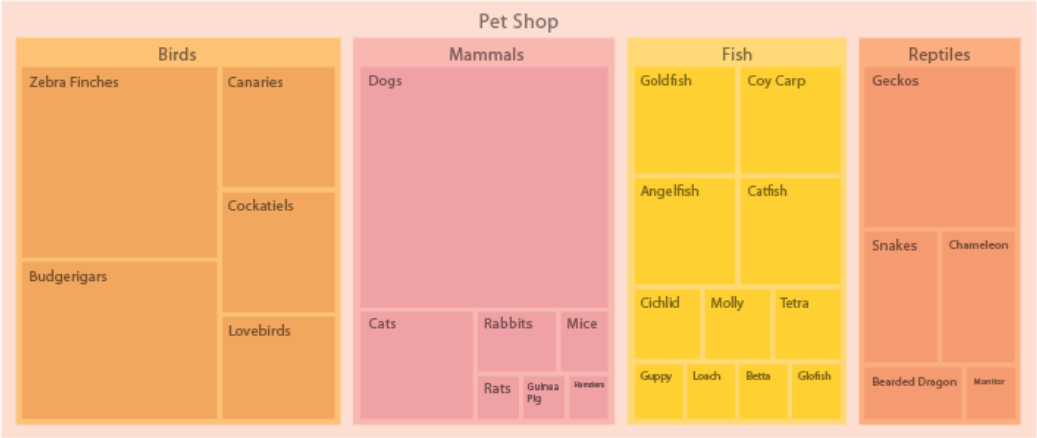



# Idioms for many categorical keys

# TreeMap: anatomy

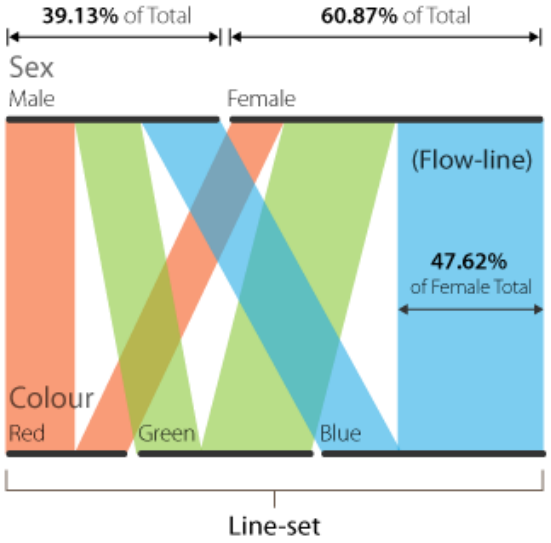
- What?
  - N categorical keys
  - 1 quantitative attribute
- Why?
  - Part-to-whole
  - Compare values
- Remarks
  - Recursion allows to manage many keys
  - Space vs Color encoding
  - Color to encode additional attribute

| class | species | num |
|---|---|---|
| Birds | Canares | 10 |
| Mammals | Dogs | 90 |
| Fish | Goldfish | 10 |
| … | … | … |



# Parallel sets: anatomy

- What?
  - N categorical attributes
- Why?
  - Correlation
  - Identify outliers, range and patterns
- Remarks
  - Scale up to hundreds of items and a few attributes
  - Color(hue) can encode additional categorical attribute

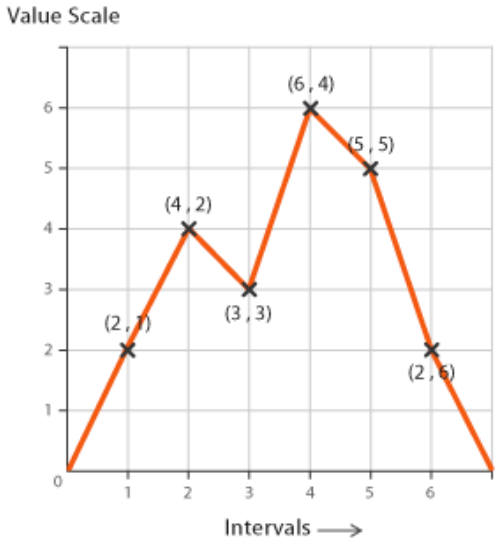


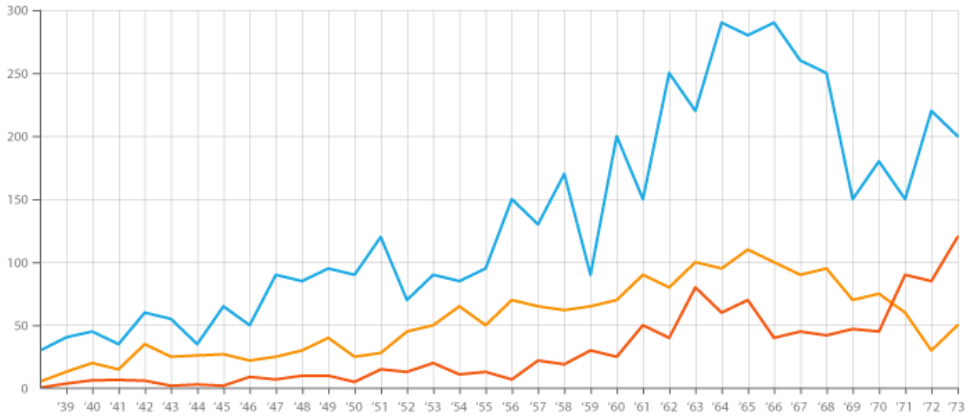

| Sex | Colour | Count | % of Sex TOTAL |
|---|---|---|---|
| Male | Red | 35 | 32.41 |
| | Green | 33 | 30.56 |
| | Blue | 40 | 37.04 |
| Female | Red | 28 | 16.67 |
| | Green | 60 | 35.71 |
| | Blue | 80 | 47.62 |

# Idioms for dealing with time

# Line graph: anatomy

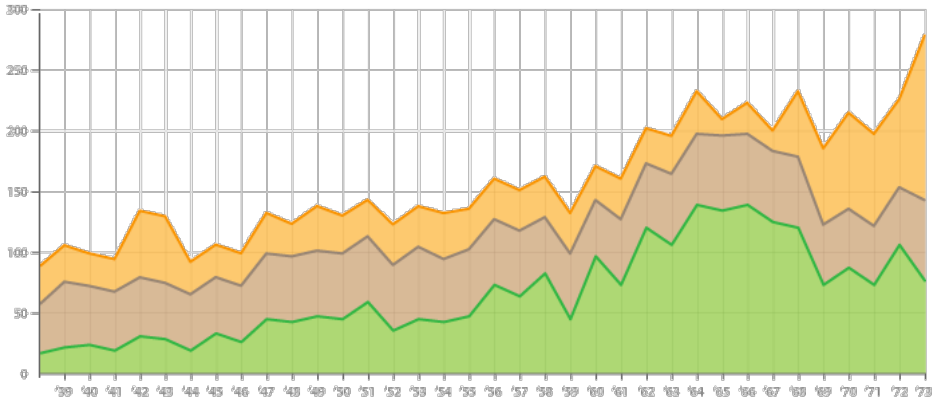
- What?
  - 1 ordered key (time)
  - 1 quantitative attribute
- Why?
  - Identify and compare trends
- Remarks
  - Scale up to 10-20 lines
  - Color can encode an additional categorical attribute



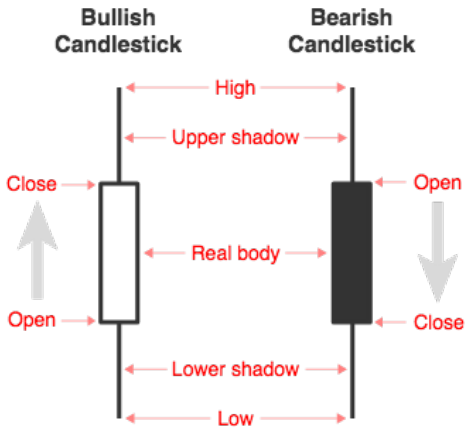
# Stacked area graph: anatomy

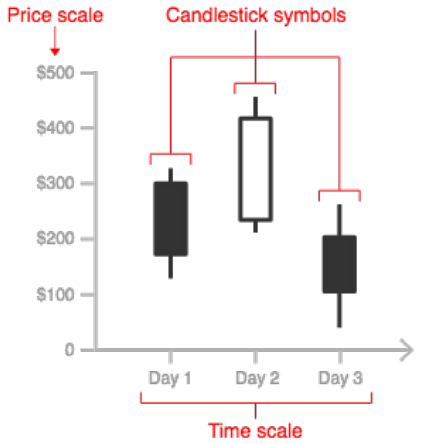
- What?
  - 1 ordered key (time)
  - 1 categorical attribute
- Why?
  - Trend
  - Part-to-whole
  - Compare values
- Remarks
  - Scale up to few values

# Candlestick chart: anatomy

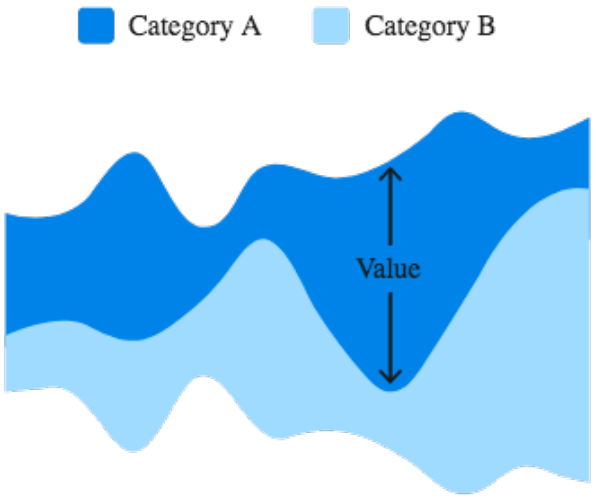
- What?
  - 1 ordered key (time)
  - 4 quantitative attributes
- Why?
  - Identify trends and patterns
- Remarks
  - Red/Green color often used instead of b/w
  - Different glyphs can be used

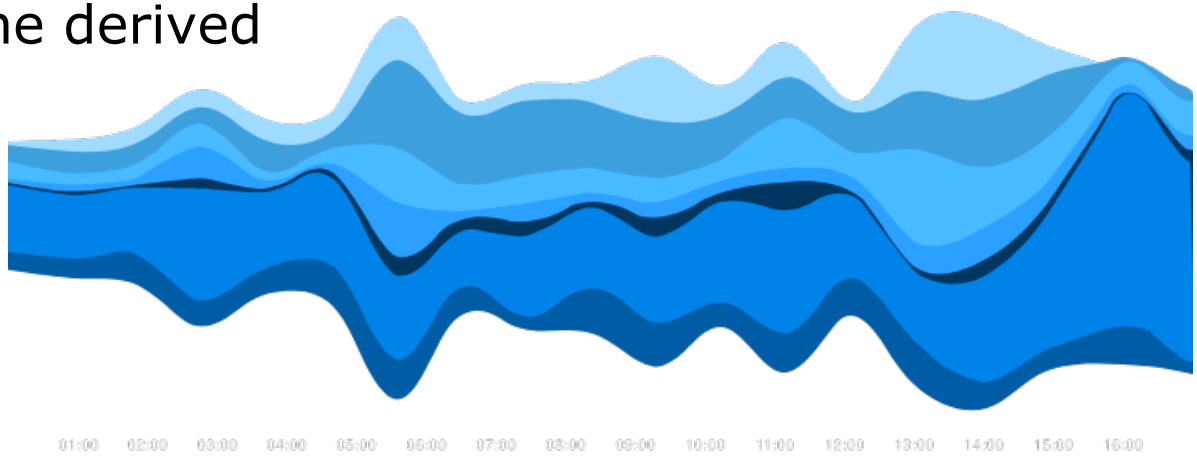


# Stream Graph: anatomy

- What?
  - 1 ordered key (time)
  - 1 categorical key
  - 1 derived attribute
- Why?
  - Trends
  - Part-to-whole
- Remarks
  - Ordering is computed to emphasize the derived attribute

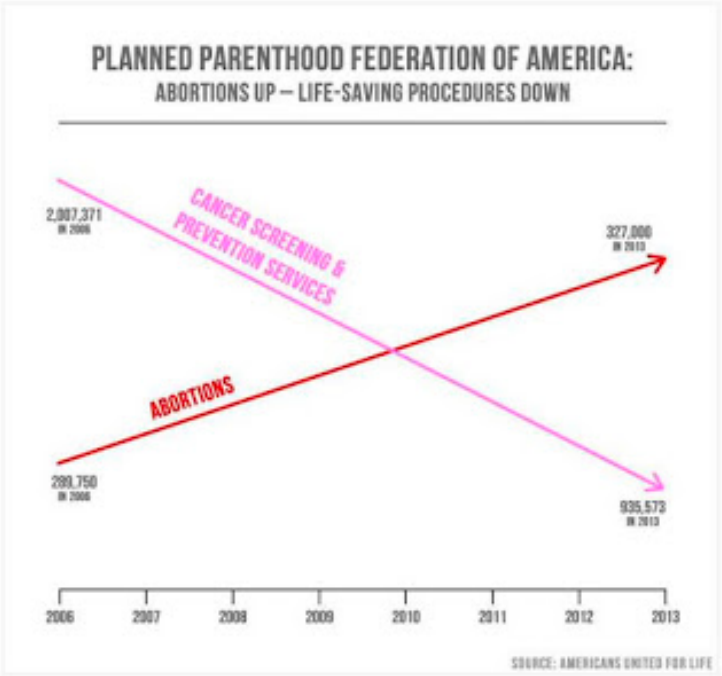


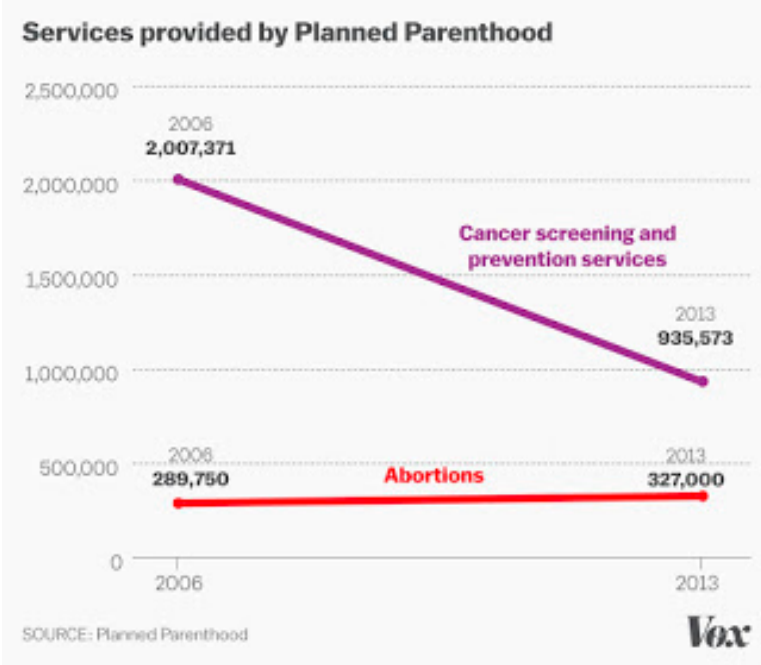



# Insights/Guidelines

# About chart axes

- Cropping y-axis:
  - Include 0 in bar charts
  - Misleading perception of slopes
- Dual axes controversial acceptable if commensurate!

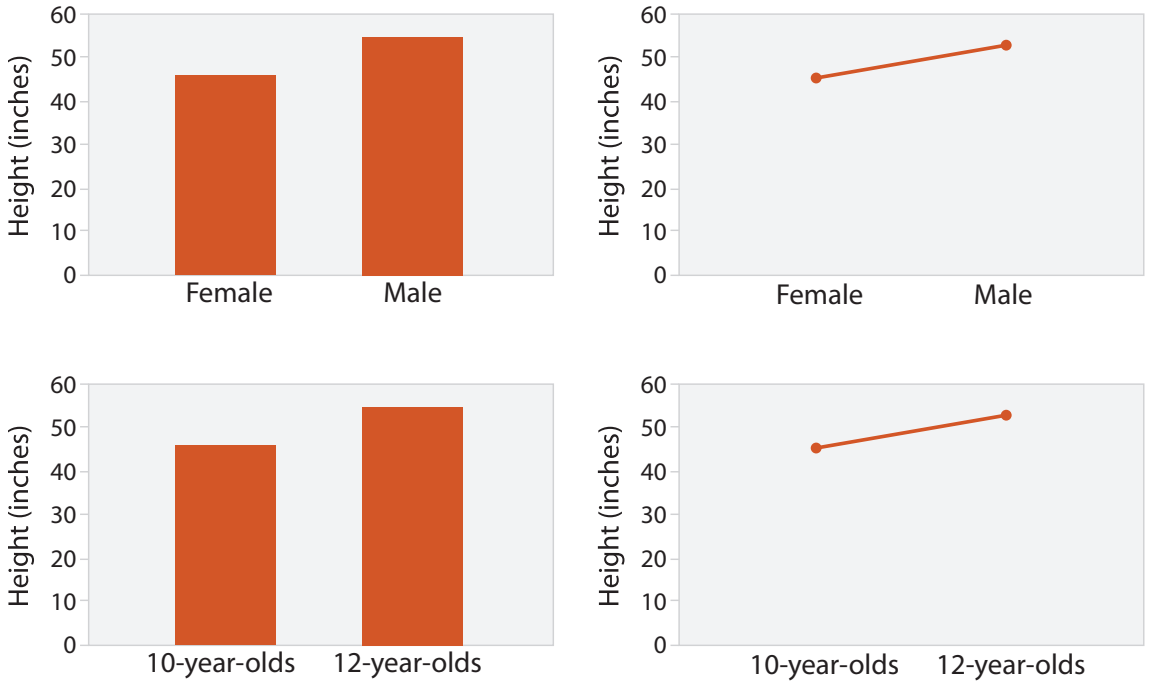



http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html

# Choosing bar vs line charts

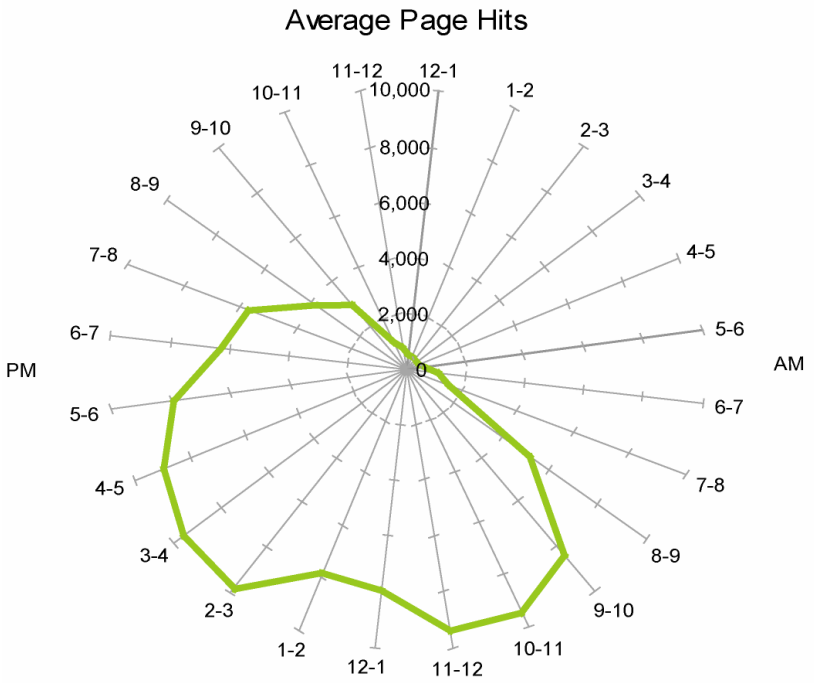
- Based on type of key attribute:
  - bar charts if categorical
  - line charts if ordered
- Implication of trend so strong that it overrides semantics!
  - "The more male a person is, the taller he/she is"




Bars and Lines: A Study of Graphic Communication. Zacks and Tversky(1999)

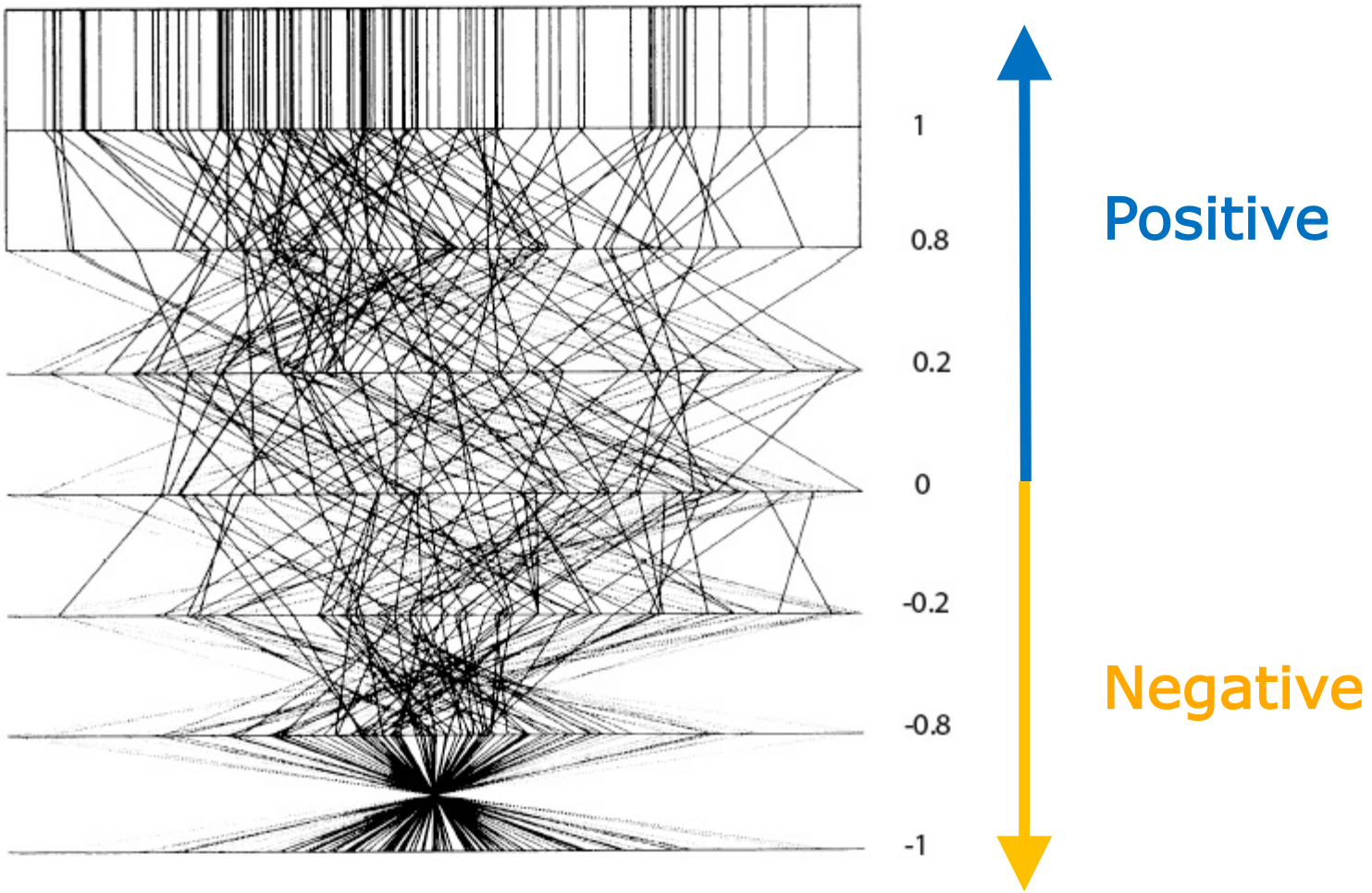
# Radial vs linear layout

- Issues with radial layout:
  - accuracy
  - efficient use of the space
- When it makes sense to use it:
  - cyclic data (e.g., months, hours, etc.)
  - data associated to radial layout (e.g., NSWE)




http://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf

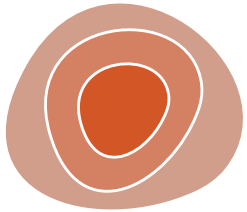
# Reading parallel coordinates

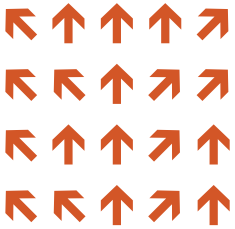


# VISUALIZING SPATIAL DATA

# Visualizing spatial data

- Geometry
  - Geographic
  - Others
- Scalar fields (one value per cell)
  - Isocontours
  - Direct volume rendering
- Vector and tensor fields (many values per cell)
  - Flow glyphs
  - Geometric (sparse seeds)
  - Textures (dense seeds)
  - Features (globally derived)

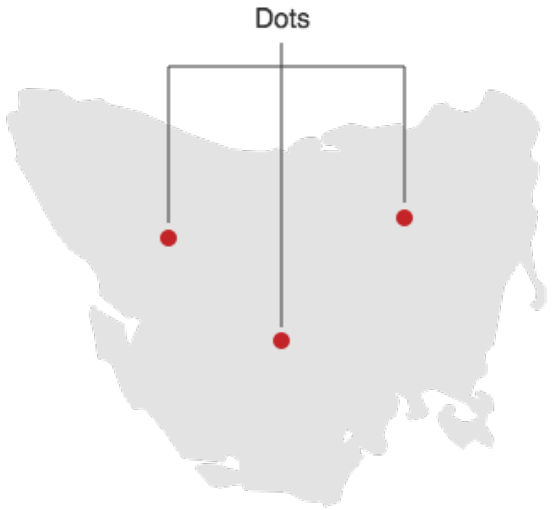


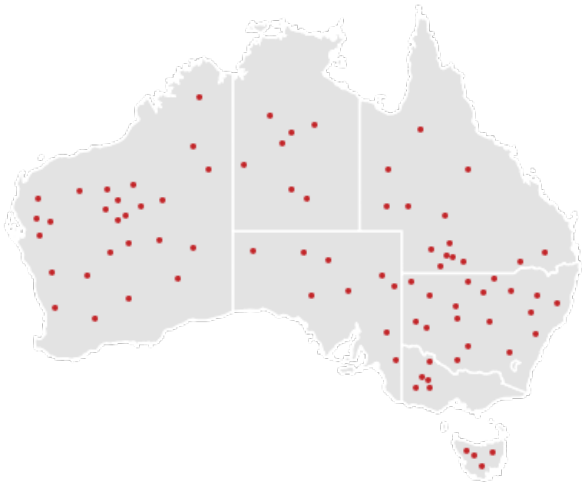


# Geometry

# Dot Map: anatomy

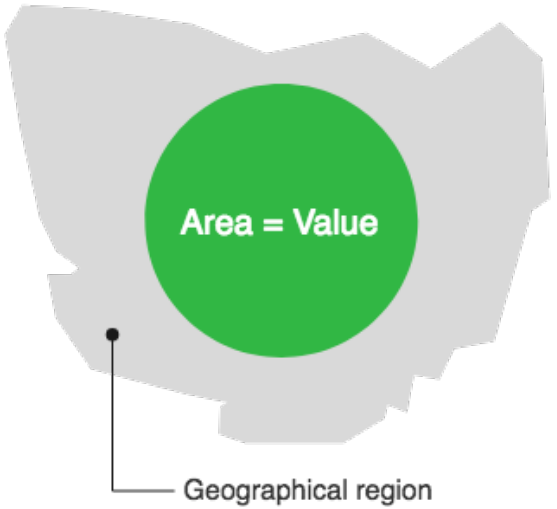
- What?
  - Geometry (position)
- Why?
  - Locate data in space
- Remarks
  - Scale up to hundreds of items
  - Color/shape can encode an additional categorical attribute (reduce scalability)

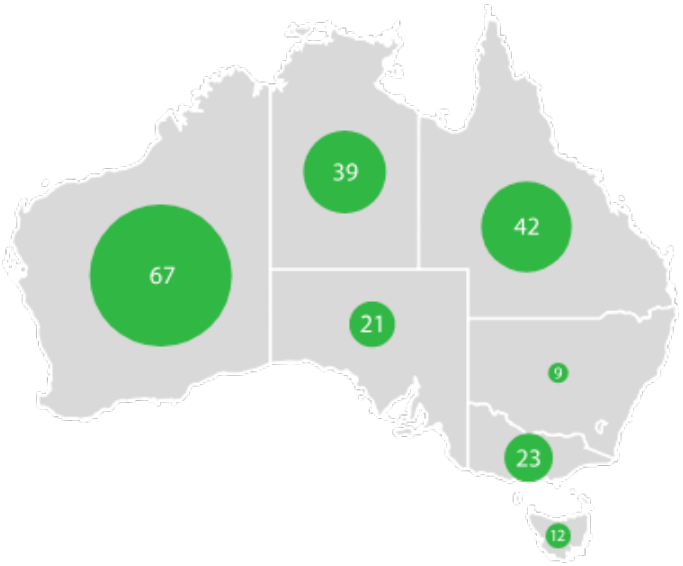


# Bubble Map: anatomy

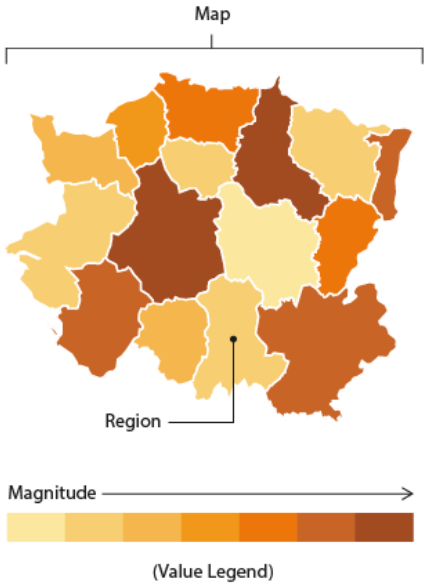
- What?
  - Geometry (position)
  - 1 quantitative attribute
- Why?
  - Locate data in space
  - Lookup and compare
- Remarks
  - Scale up to hundreds of items
  - Color can encode an additional categorical attribute (interaction with size).

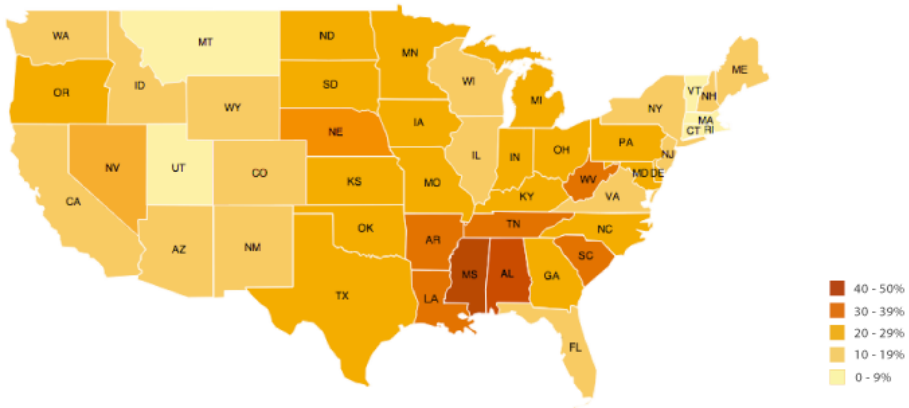


# Choropleth map: anatomy

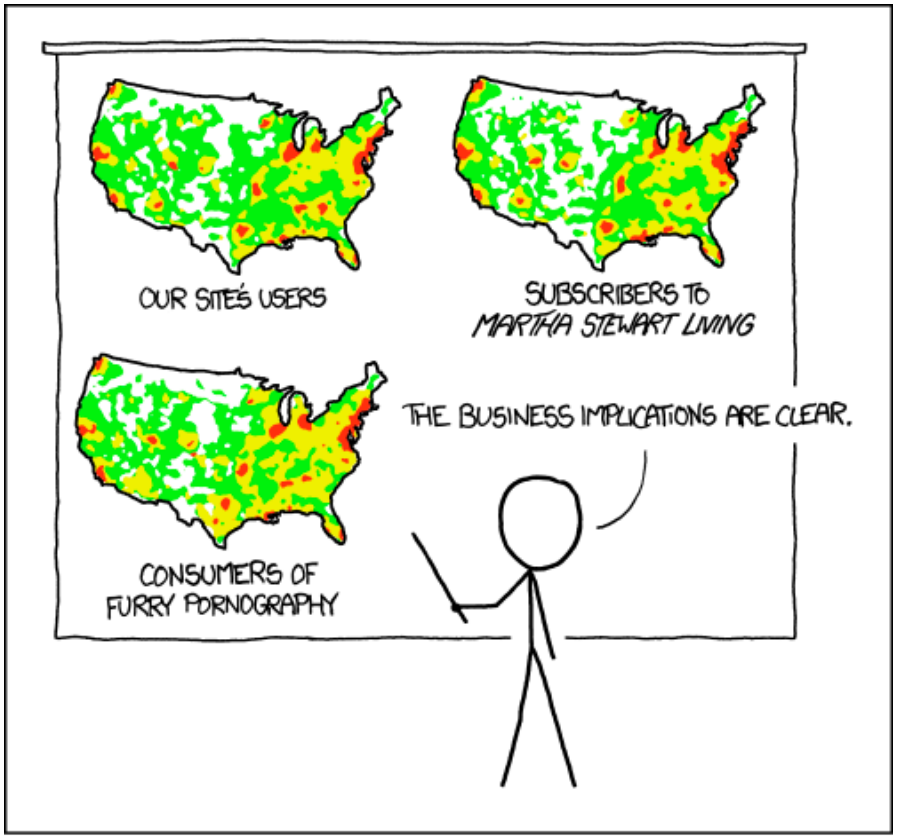
- What?
  - Geometry (position)
  - 1 quantitative attribute
- Why?
  - Locate data in space
  - Lookup and compare
- Remarks
  - Scale up to ~1000 items
  - Hue can encode an additional categorical attribute (better if binary)



# Bias of population-related maps

- Using absolute value is dangerous: any map would look like the populatio map!
- How to deal with this?
  - Visualize per capita (relative) values
  - Use statistical models

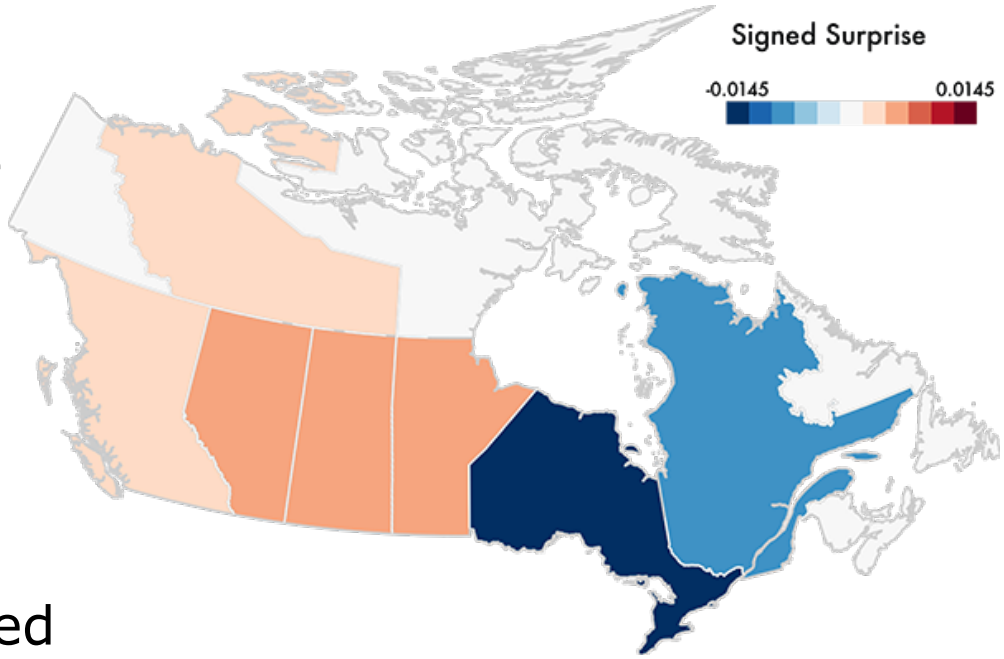



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

[ https://xkcd.com/1138 ]

# Surprise map: anatomy

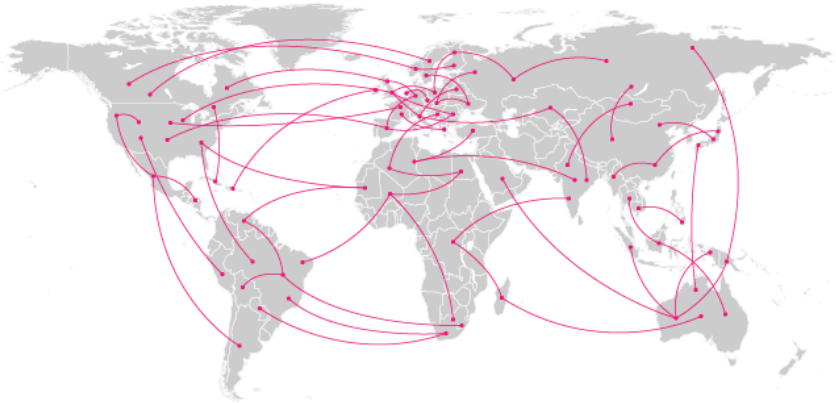
- What?
  - Geometry (position)
  - 1 quantitative attribute
  - 1 derivative attribute
- Why?
  - Locate data in space
  - Lookup and compare
- Remarks
  - The surprise is computed as a function prior and posterior probability of data distribution.
  - Prior probability is generated with a family of standard models.

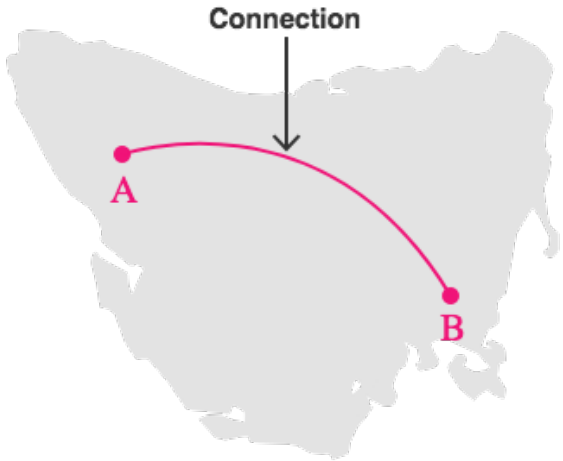



Correll & Heer (2016)

# Connection map: anatomy

- What?
  - Network and positions
- Why?
  - Lookup path
  - Identify patterns
- Remarks
  - Size of links can encode an additional ordered attribute (3-4 bins at max)

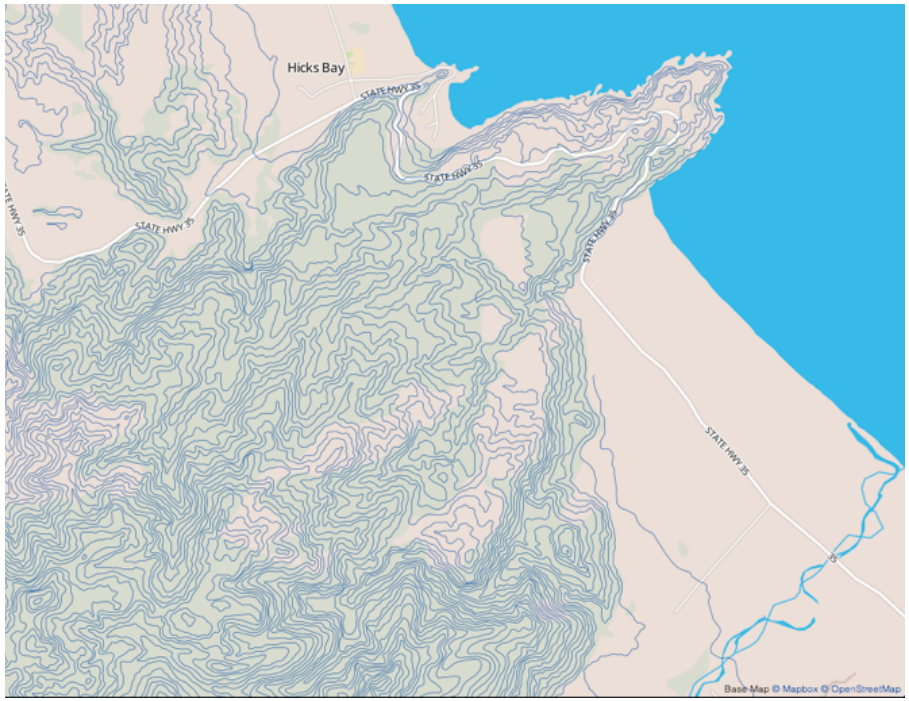




# Scalar Fields

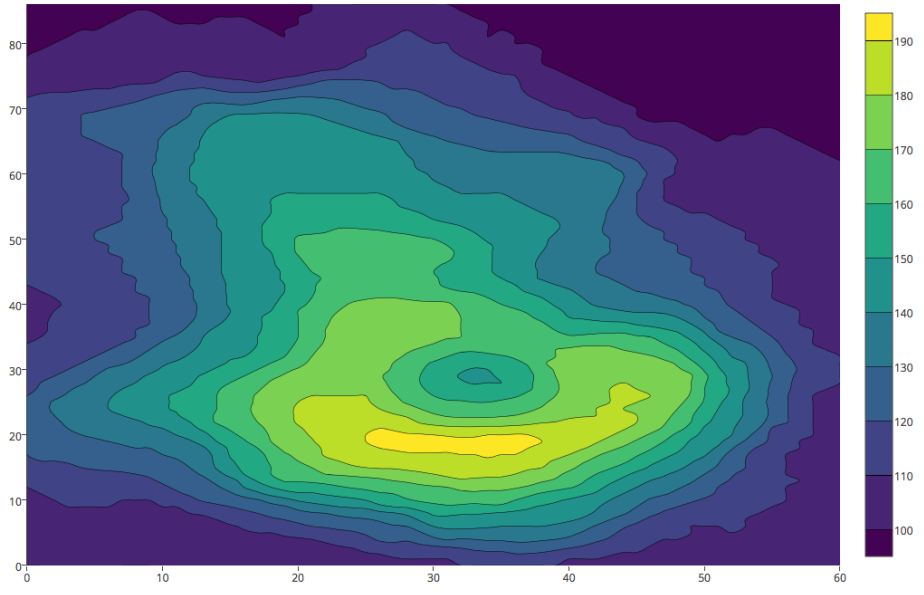
# Isocontour map: anatomy

- What?
  - Geographic data
  - 1 quantitative attribute
  - derived positions
- Why?
  - Shape
- Remarks
  - The lines are computed from the values of scalar field
  - Area can be filled and color encoded



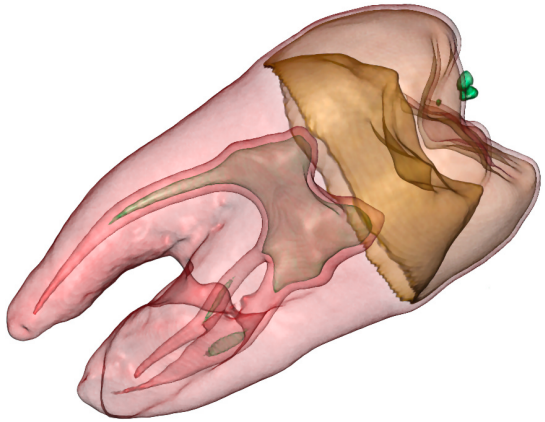
# Isocontour plot: anatomy

- What?
  - 2D spatial field
  - 1 quantitative attribute
  - derived geometry
- Why?
  - Shape and patterns
- Remarks
  - The lines are computed from the values of scalar field
  - Area can be empty or filled and color encoded



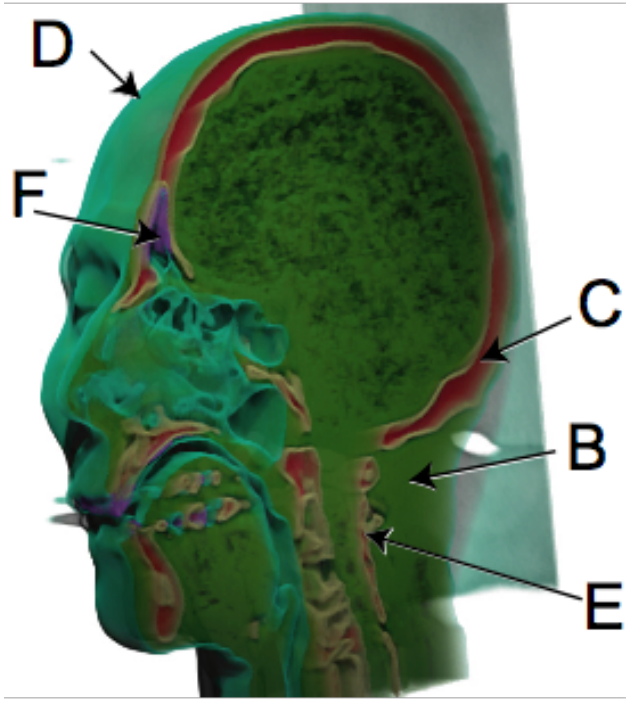
# Isosurface plot: anatomy

- What?
  - 3D spatial scalar field
  - 1 quantitative attribute
  - derived geometry
- Why?
  - Shape
- Remarks
  - Tree of isosurfaces: positions computed for specific values of the scalar field



# Direct volume rendering: anatomy

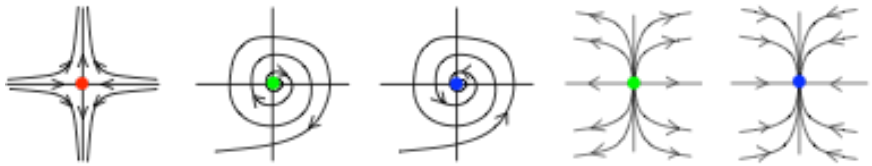
- What?
  - 3D spatial scalar field
  - 1 quantitative attribute
- Why?
  - Shape
- Remarks
  - The values in the field (and the gradient) are mapped directly to color and opacity

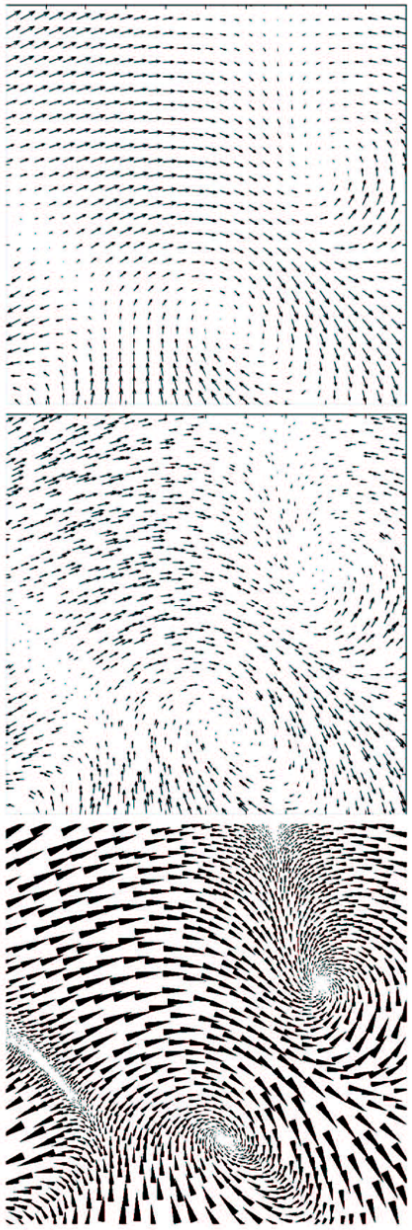


# Vector field

# Glyph flow: anatomy

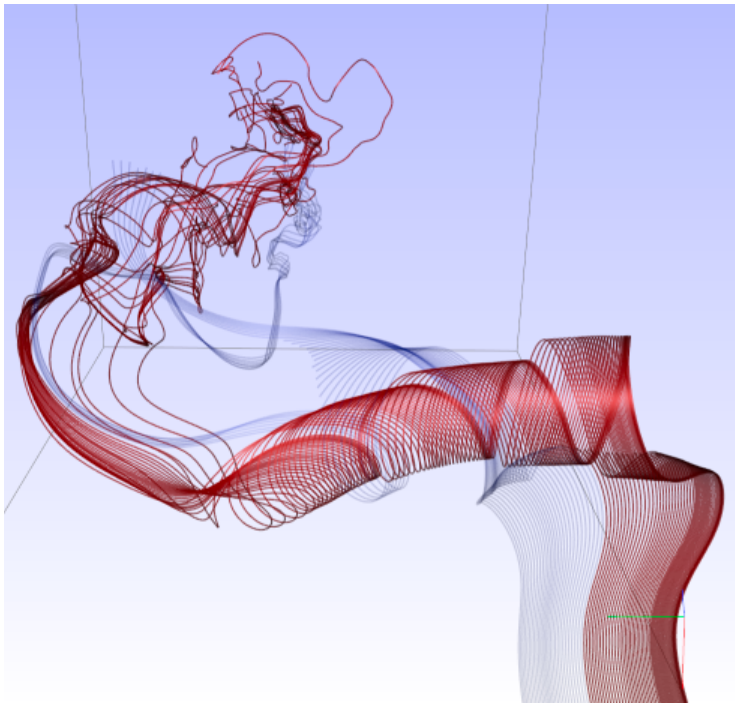
- What?
  - 2D vectorial field
- Why?
  - Shape and patterns
  - Identify critical points
- Remarks
  - Different glyphs can be used to represent vectors
  - Density of grid and jittering





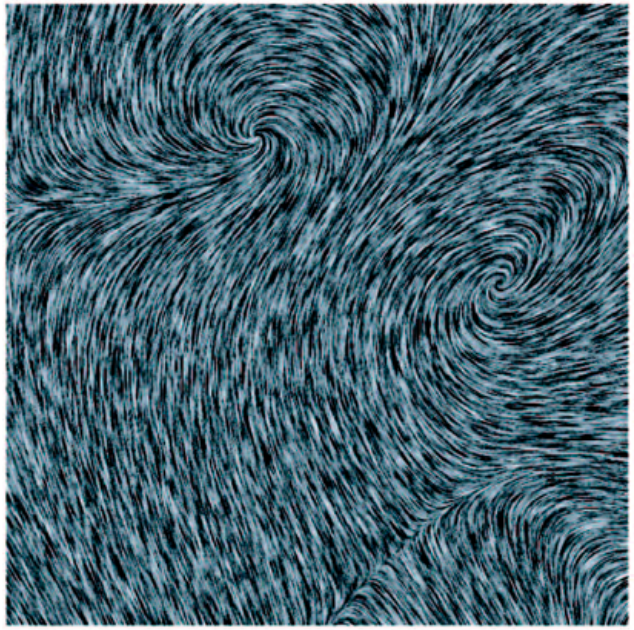
# Geometric flow: anatomy

- What?
  - 2D/3D vectorial field
  - derived geometry
- Why?
  - Shape and patterns
  - Identify critical points
- Remarks
  - Seeding strategy affects the outcome
  - Usage of clustering and color coding improves readability



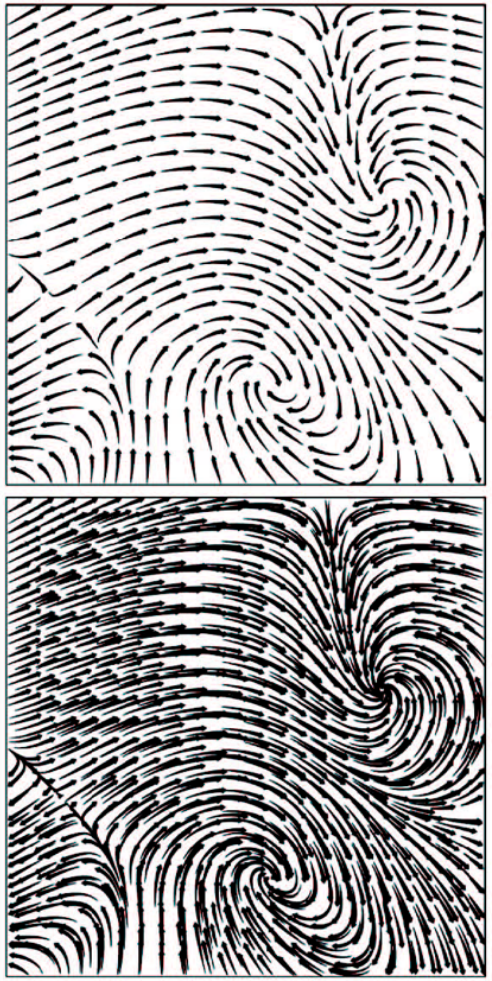
# Texture flow: anatomy

- What?
  - 2D vectorial field
- Why?
  - Shape and patterns
  - Identify critical points
- Remarks
  - Similar to glyphs flow, but computes the flow of a continuous distribution of particles



# Feature flow: anatomy

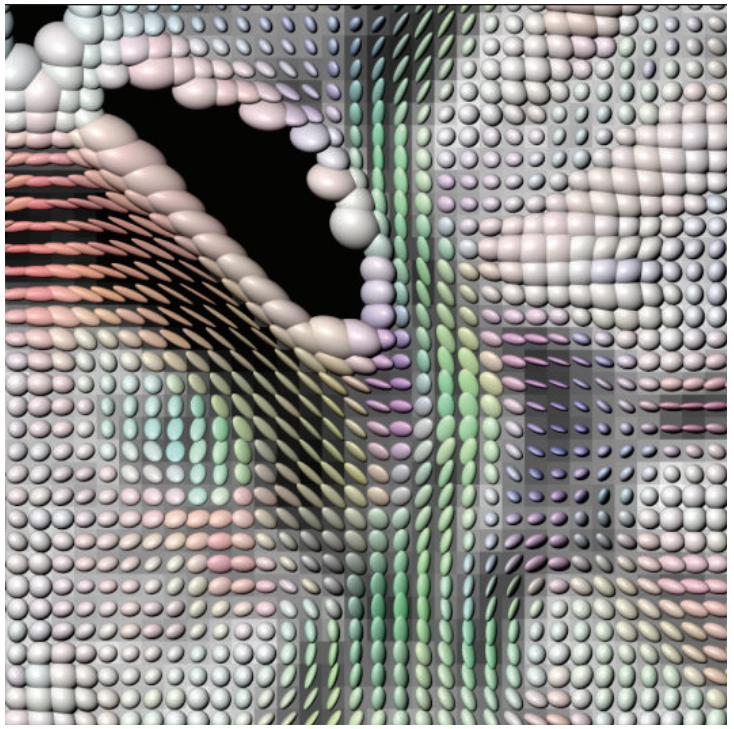
- What?
  - 2D vectorial field
- Why?
  - Shape and patterns
  - Identify critical points
- Remarks
  - Similar to glyphs flow, but seeding is based on global computing strategy to identify areas with similar behaviors



# Tensor field

# Ellipsoid tensor glyphs: anatomy

- What?
  - 2D/3D tensor field
- Why?
  - Shape and patterns
- Remarks
  - Use shape, orientation, color, opacity to represents field tensors



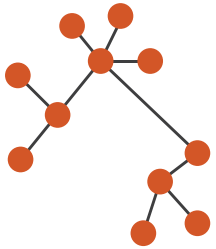
# VISUALIZING TREES AND NETWORKS

# Major approaches

## Node–Link Diagrams

Connection Marks

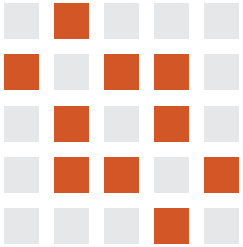
✔ NETWORKS ✔ TREES



## Adjacency Matrix

Derived Table

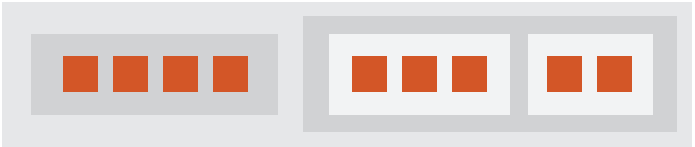
✔ NETWORKS ✔ TREES



## Enclosure

Containment Marks

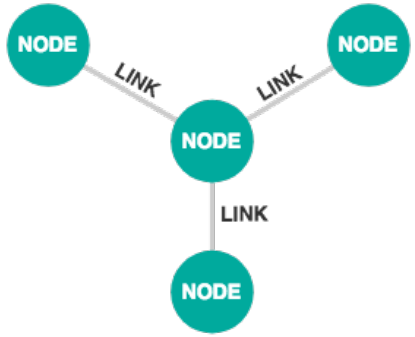
✖ NETWORKS ✔ TREES

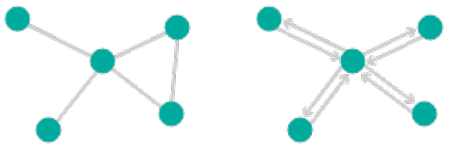


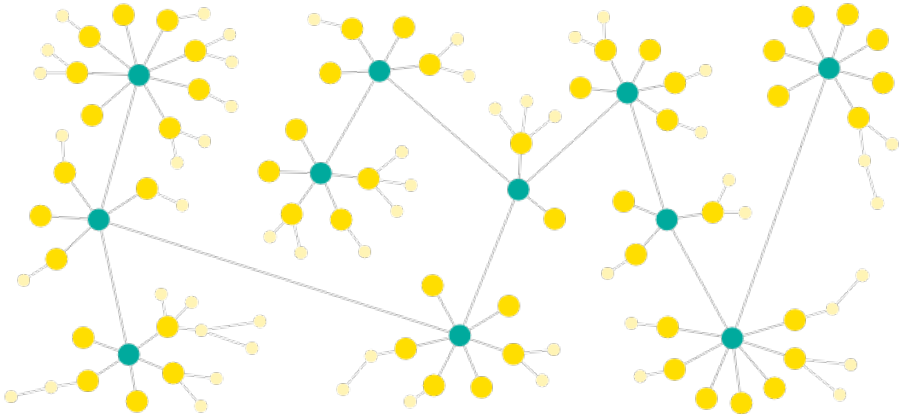
# Node-Link Diagrams

# Node-link diagram: anatomy

- What?
  - Network
- Why?
  - Explore topology and locate paths
- Remarks
  - Different algorithms to compute layout (sfdp, planar, radial, etc.)
  - **Hairball effect**

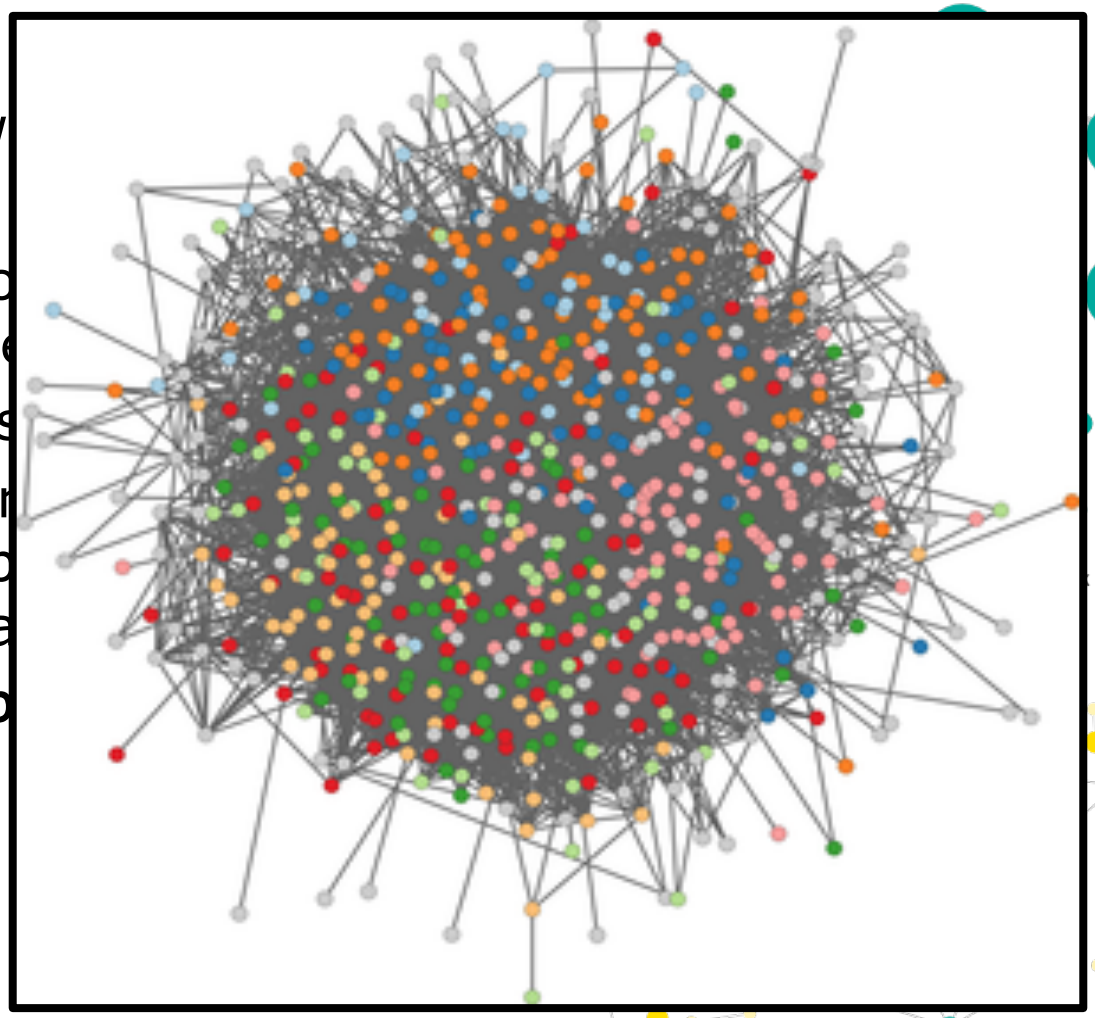


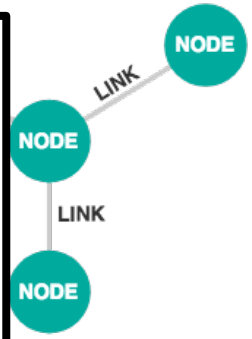



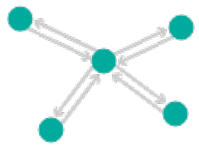



# Node-link diagram: anatomy

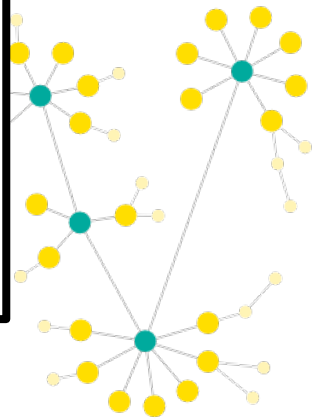
- What?
  - Netw
- Why?
  - Explo
    locate
- Remarks
  - Differ
    comp
    plana
  - **Hairb**





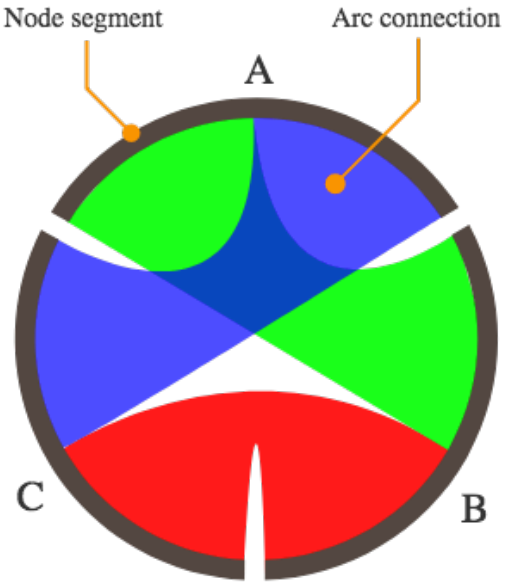



Directed network



# Chord diagram: anatomy

- What?
  - Network (adjacency matrix)
- Why?
  - Explore connections
  - Identify patterns
- Remarks
  - Size is used to encode quantitative attribute of links
  - Color encodes an additional categorical attribute (or grouping)




| | A | B | C |
|---|---|---|---|
| A | | 10 | 10 |
| B | 10 | | 10 |
| C | 10 | 10 | |

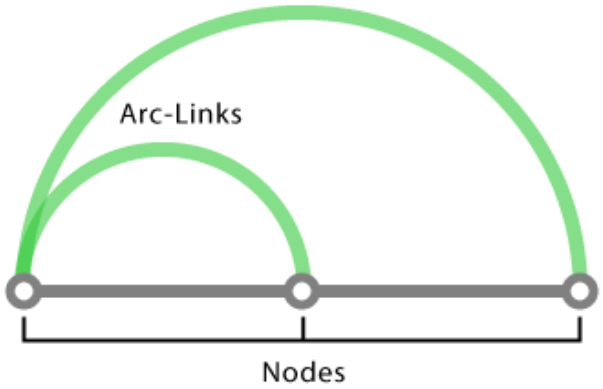
A $\xrightarrow{10}$ B

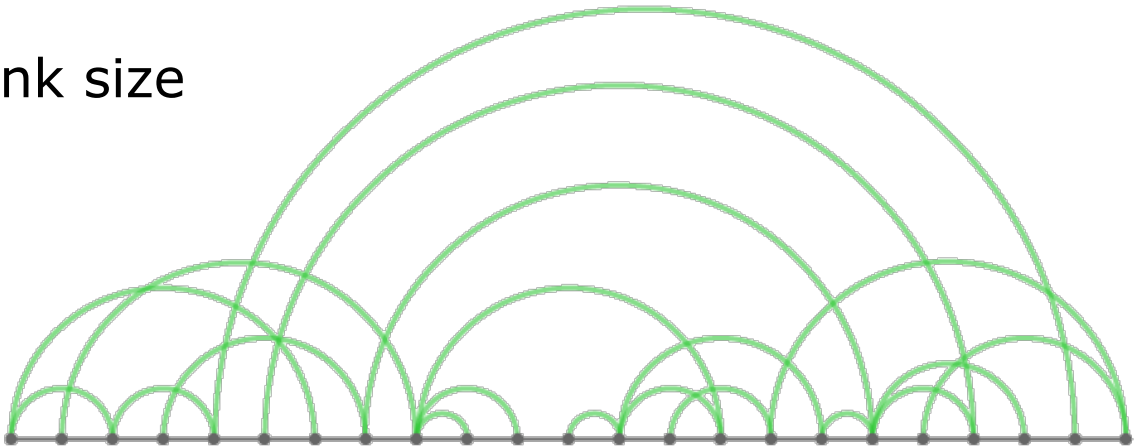
B $\xrightarrow{10}$ C

C $\xrightarrow{10}$ A

# Arc diagram: anatomy

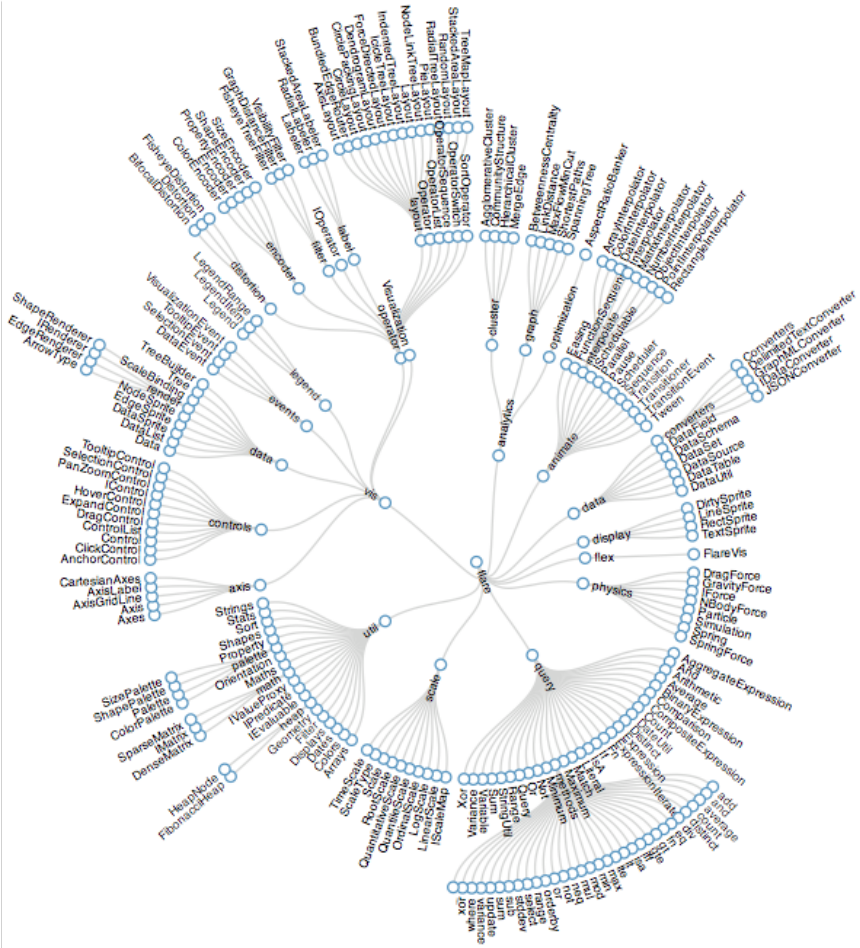
- What?
  - Network (links)
- Why?
  - Locate links and paths
  - Identify patterns
- Remarks
  - Don’t show the topology
  - Encode additional attribute of with link size (ordered) or color (categorical)



# Node-link diagram: anatomy

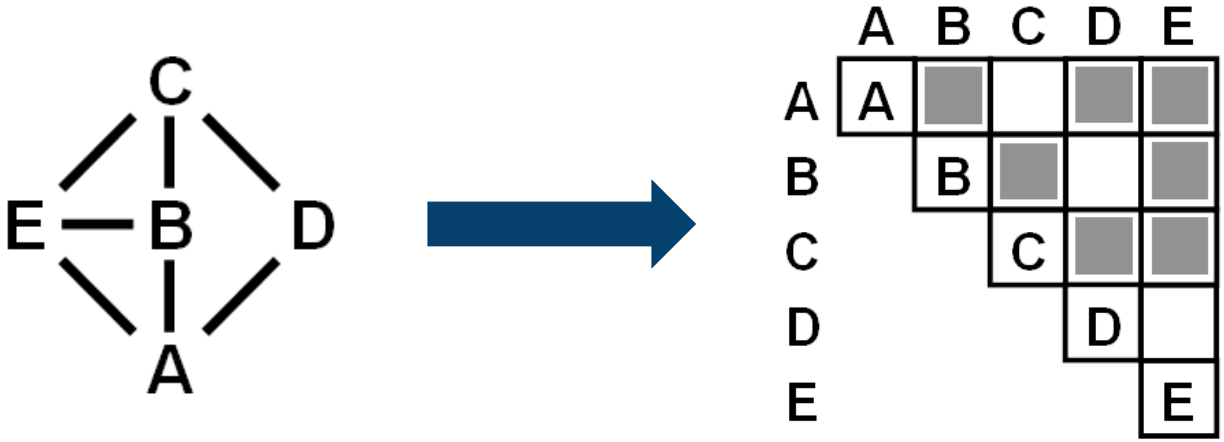
- What?
  - Tree
- Why?
  - Explore topology and locate paths
- Remarks
  - Scale up to 1-10K nodes



# Adjacency Matrix

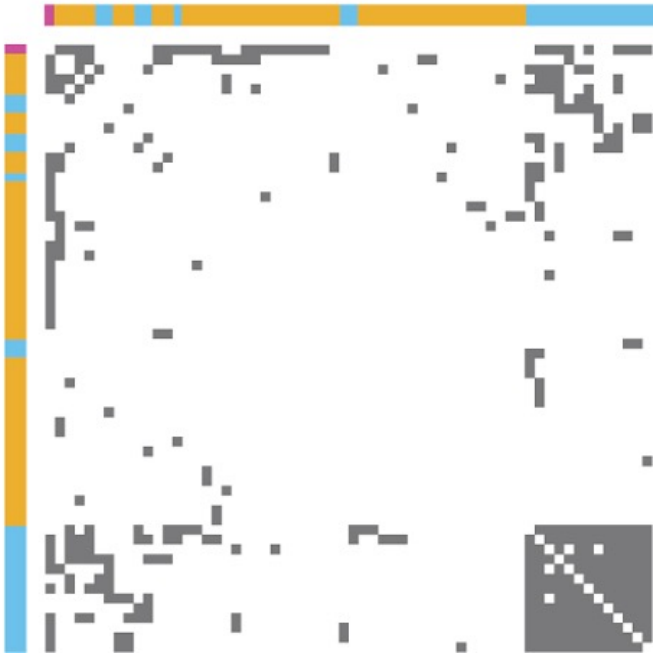
# Representing networks with matrix

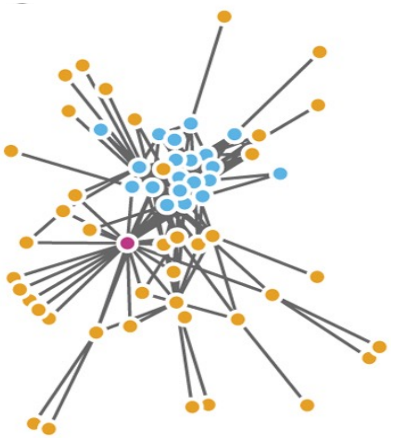
- Adjacency matrix are built from a network:
  - size NxN, with N is the number of nodes
  - M[i,j]≠0 if node i and j are connected
  - M[i,j] might also encode a quantitative attribute of the link



# Adjacency matrix: anatomy

- What?
  - Network
  - Derived Table (2 categorical, 1 quantitative)
- Why?
  - Discover cluster and connectivity patterns
- Remarks
  - Scale up to 1000 nodes and 1M links
  - Color can encode a quantitative attribute of links (e.g., weight)

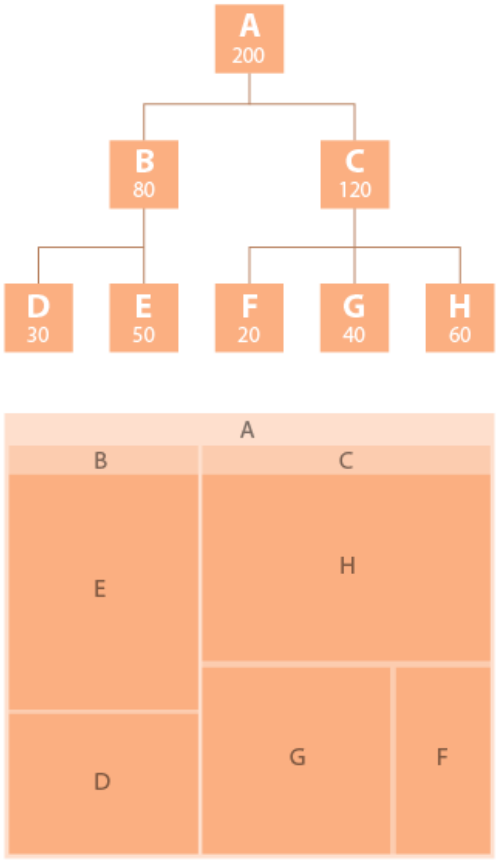




# Enclosure

# TreeMap: anatomy

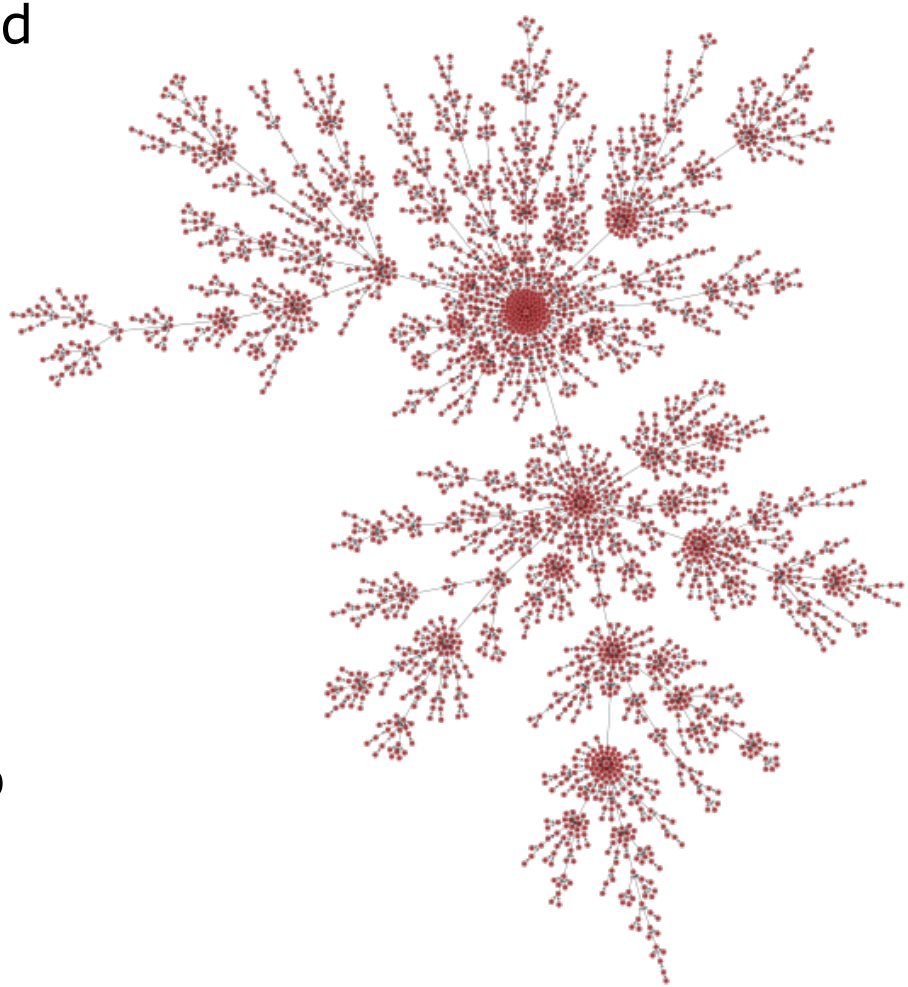
- What?
  - Tree
- Why?
  - Part-to-whole
  - Identify hierarchy
- Remarks
  - Recursion allows to manage many levels
  - Color to encode an additional attribute



# Insights

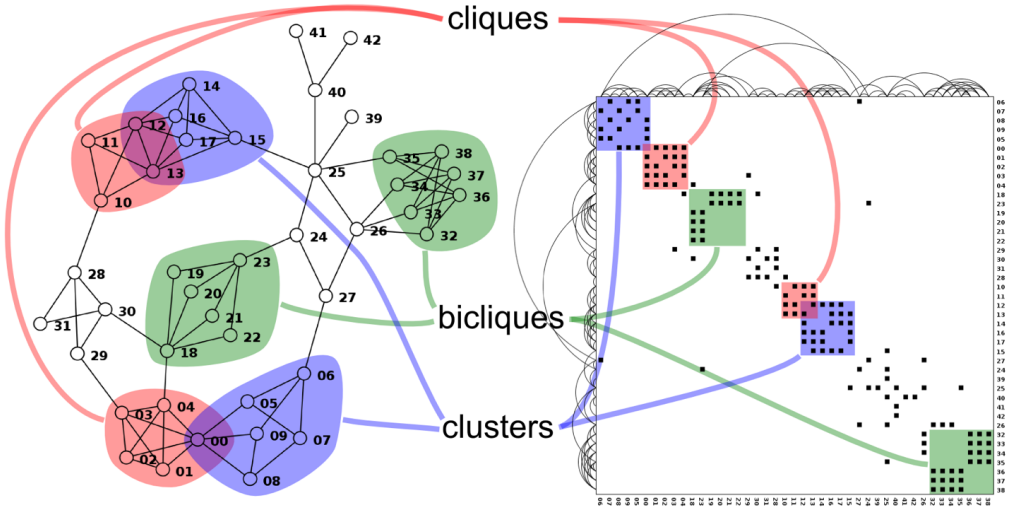
# Force-Directed Placement

- A popular algorithm to build network layout based on a simulation of physical forces:
  - Node repel each others
  - Links act as springs
  - Scales up to 50-100 nodes and links < 4*nodes
- A multilevel version (sfdp) try to preserve local structures and scales up to 1K-10K nodes and links < 4*nodes



# Adj matrix vs node-link diagrams

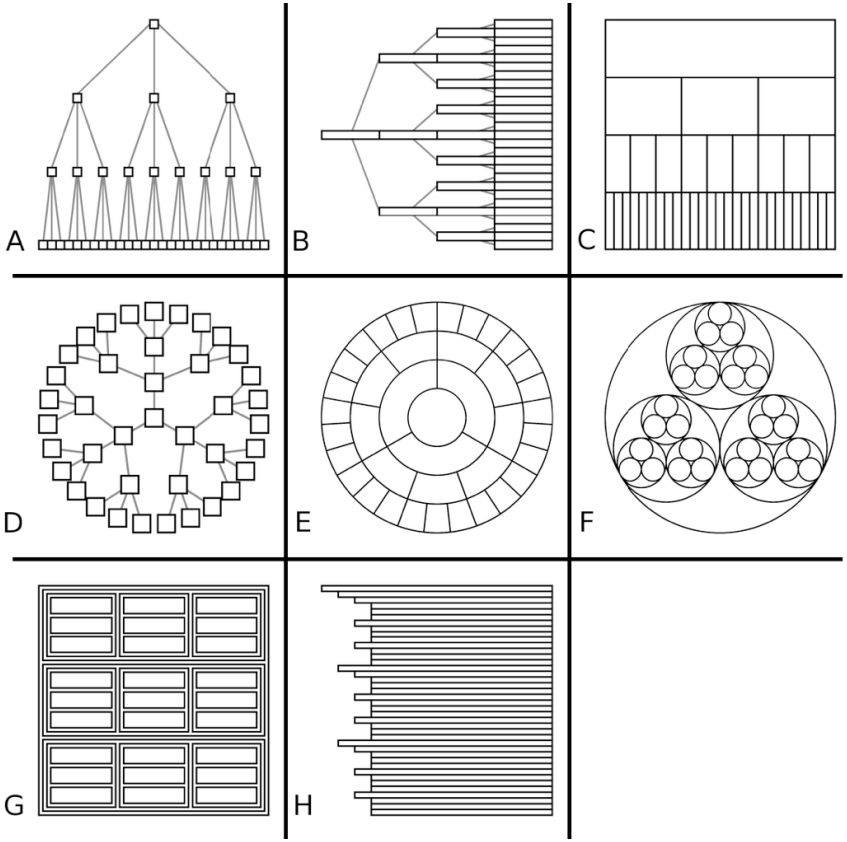
- Benefit of adj matrix:
  - predictability, scalability, supports reordering
  - some topologies can be identified with training
- Benefit of node-link diagram:
  - topology understanding, path tracing
  - intuitive, no training needed




- Guidelines
  - node-link best for small networks
  - matrix best for large networks (if topology is not the focus)

# Layouts for trees

- Data to visualize:
  - link relationships
  - tree depth
  - sibling order
- Design choices
  - rectilinear or radial layout
  - connection or containment




[See McGuffin and Robert (2010)]