

# Visualization Techniques

Pere-Pau Vázquez

ViRVIG Group

Dept. Computer Science – UPC

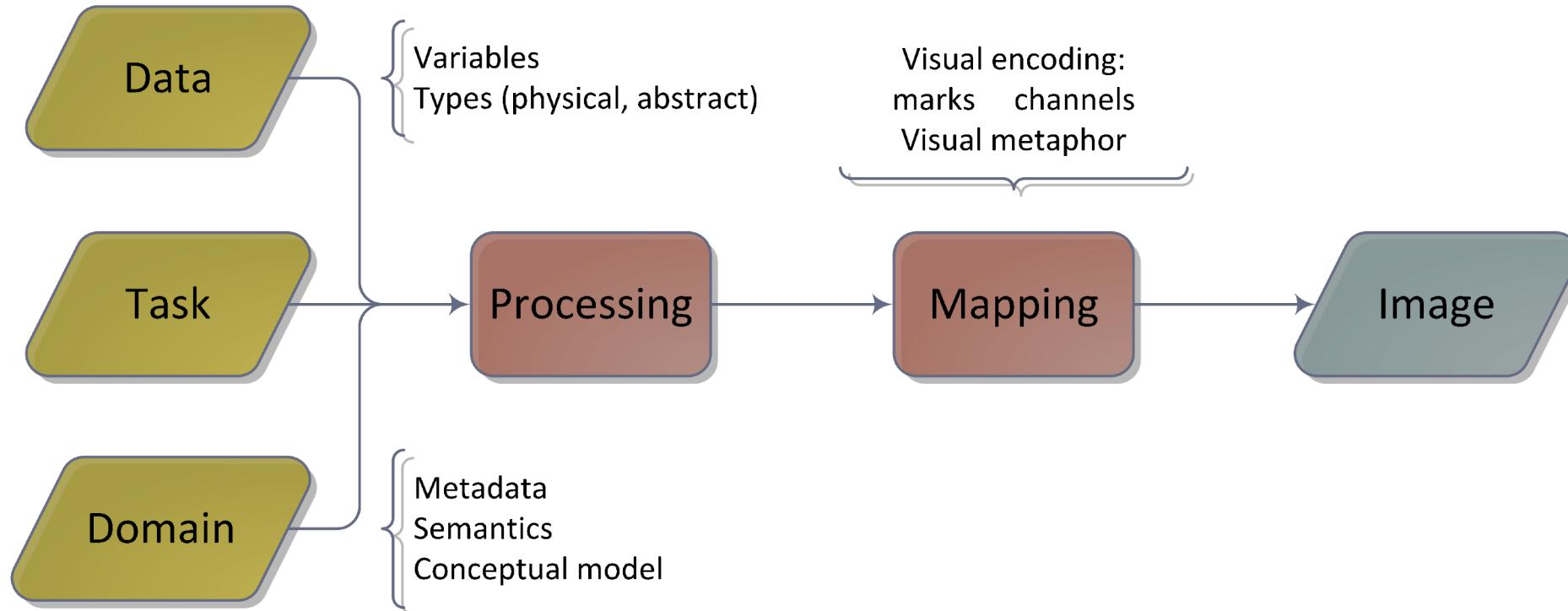
# Outline

- Basics
- Basic data representations
- Advanced data representations
- Multi-functioning elements
- Further reading

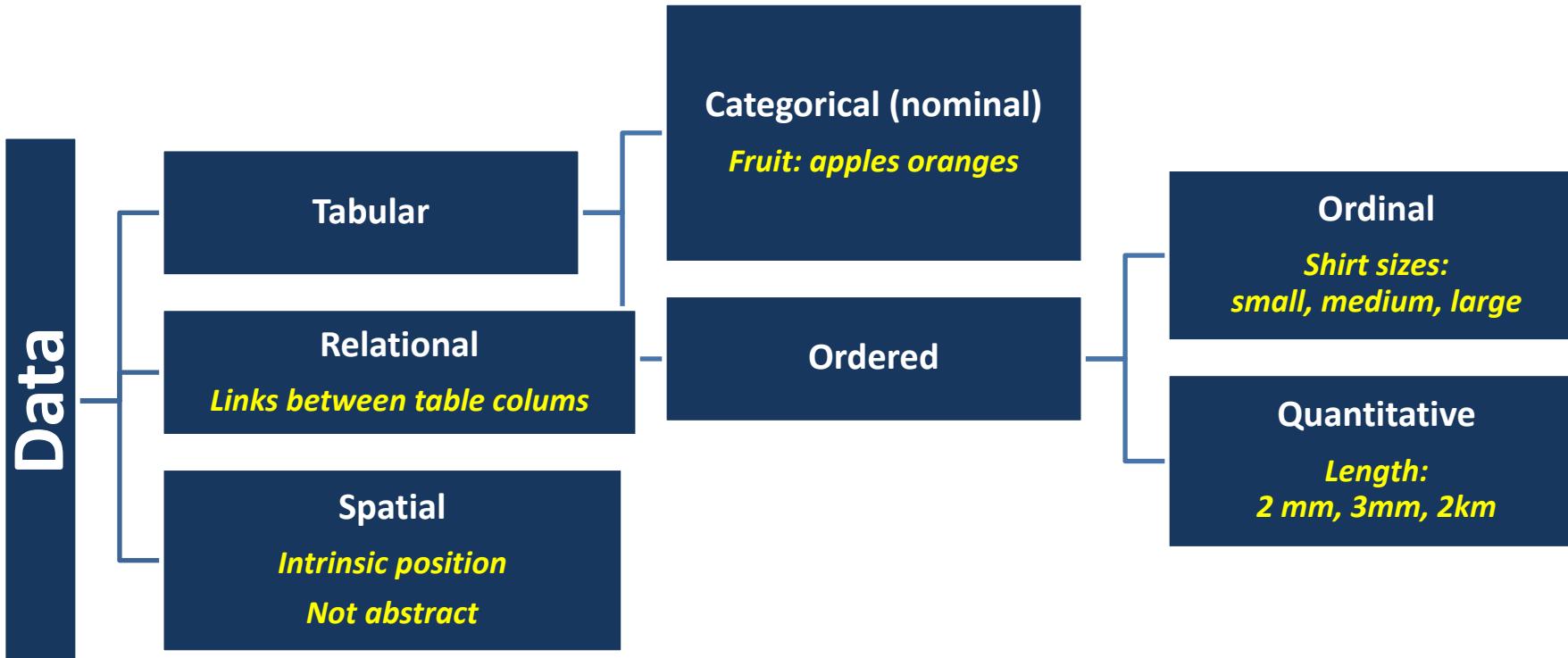
# Basics

Pere-Pau Vázquez  
Dept. Computer Science – UPC

# Recap. Visualization process



# Recap. Data



# Recap. Visual encodings

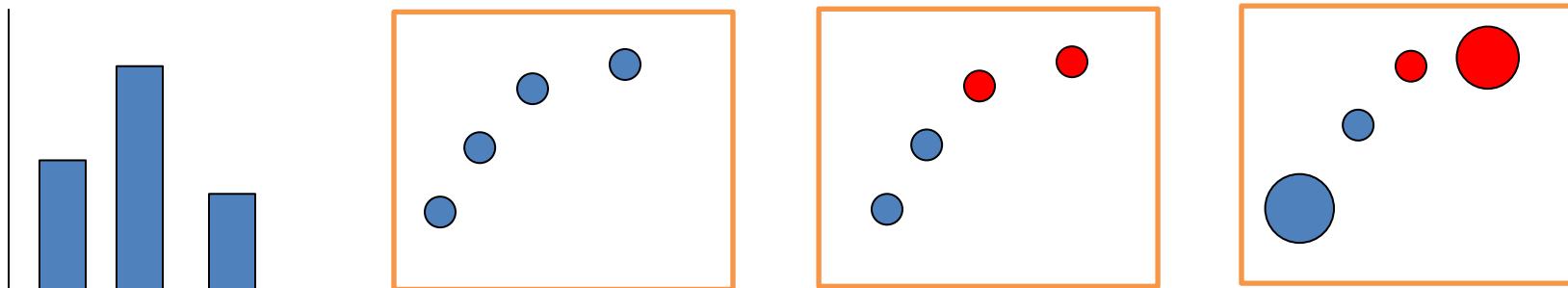
- Huge space of design alternatives
  - Tradeoffs abound
- Many possibilities now known to be *ineffective*
  - Avoid random walk through the possibilities
  - Avoid some of known mistakes
  - Extensive experimentation has already been done
- Guidelines continue to evolve
  - We reflect on lessons learned in design studies
  - Iterative refinement usually wise

# Recap. Visual encodings

- Know your visual channel types and ranks
- Color constraints
- Power of the plane, danger of depth (3D)
- Resolution beats immersions
- Eyes beat memory

# Recap. Visual encodings

- Analyze as combination of marks and channels showing abstract data dimensions



# Recap. Marks & Visual channels

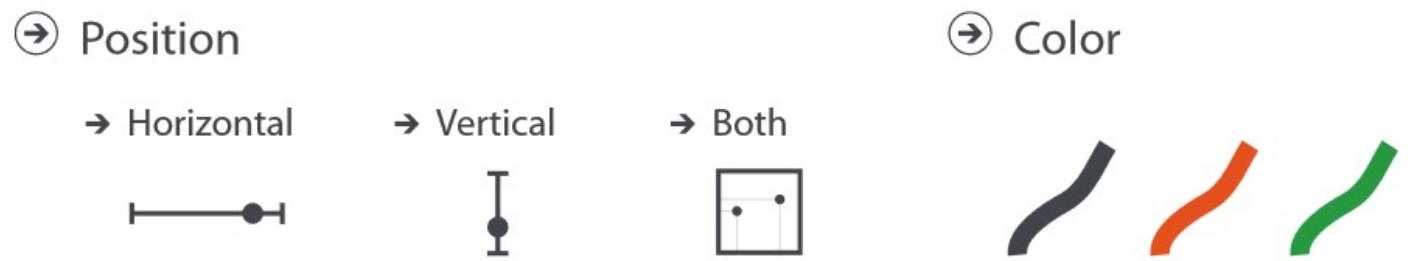
- Marks:
  - Geometric primitives
- Visual channels: control appearance of marks
  - Control appearance of marks
  - Can redundantly code with multiple channels

# Recap. Marks & Visual channels

- Marks

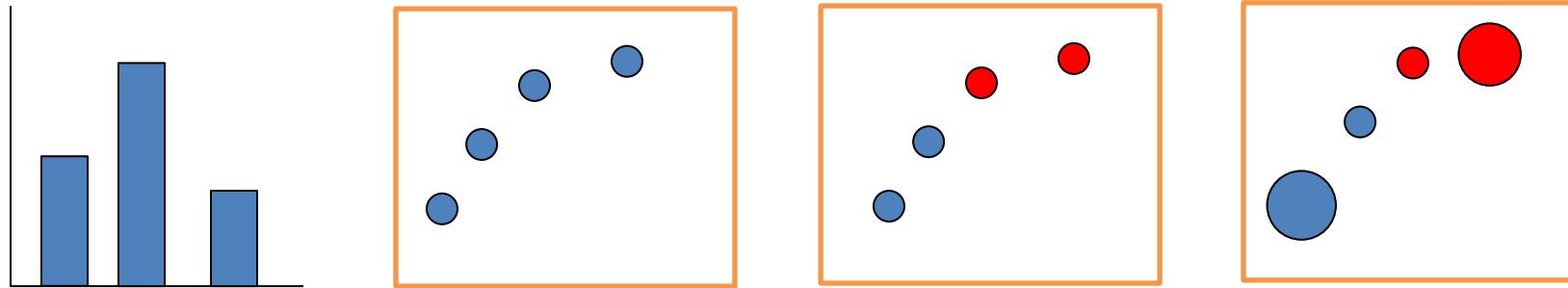


- Visual channels



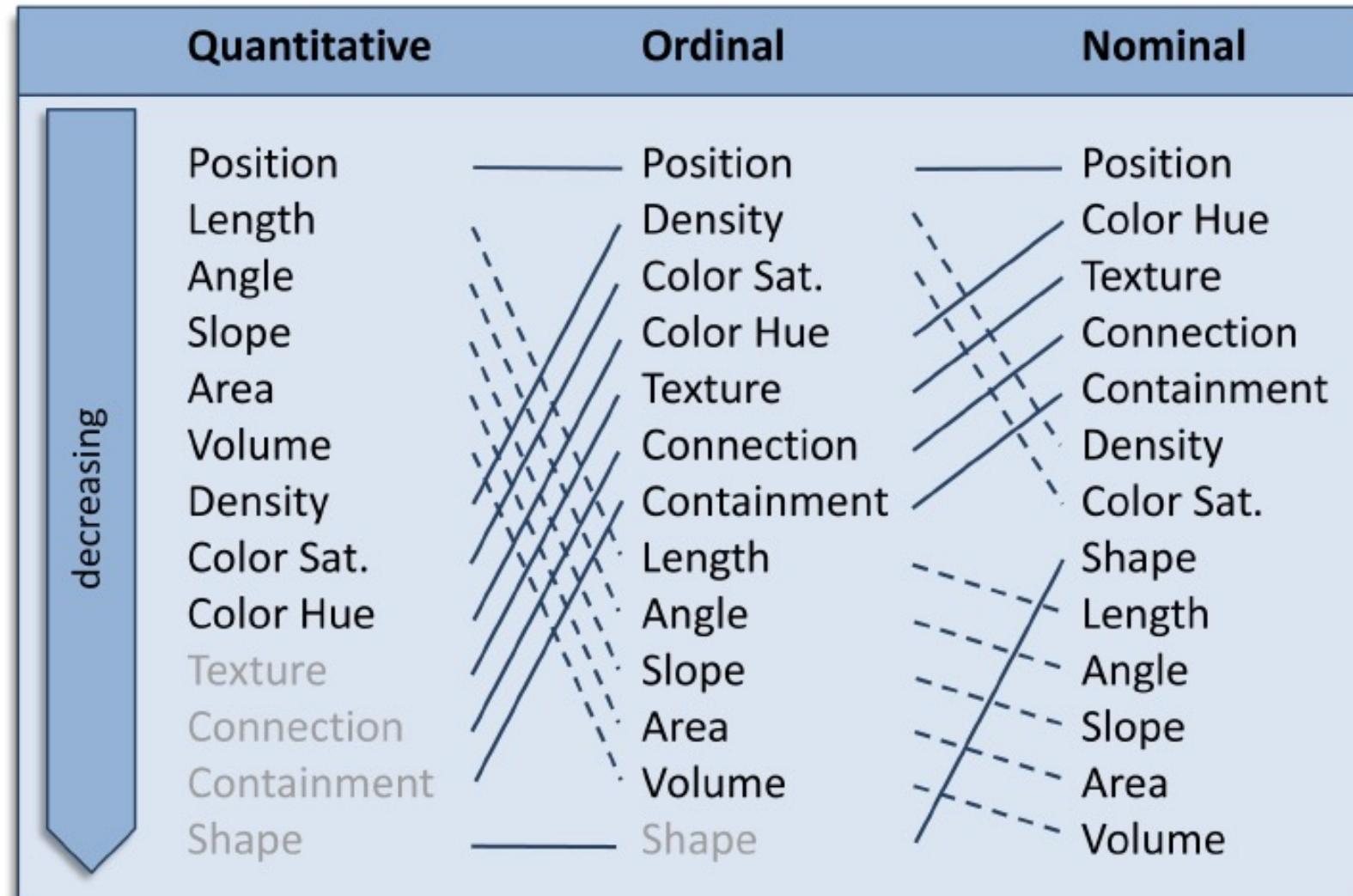
# Recap. Visual Encoding

- Analyze as combination of marks and channels showing abstract data dimensions



Channel	Vertical pos.	Vertical pos. Horizontal pos.	Vertical pos. Horizontal pos. Color	Vertical pos. Horizontal pos. Color Size
Mark	Line	Point	Point	point

# Recap. Visual Encoding. Channel ranking



# Recap. Visual Encoding. Channel ranking

## → **Magnitude Channels: Ordered Attributes**

## Position on common scale



## Position on unaligned scale



## Length (1D size)



## Tilt/angle



## Area (2D size)



### Depth (3D position)



## Color luminance



## Color saturation



## Curvature



### Volume (3D size)



## → Identity Channels: Categorical Attributes

## Spatial region



## Color hue



## Motion



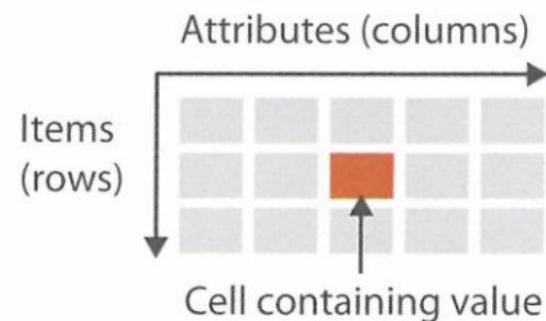
## Shape



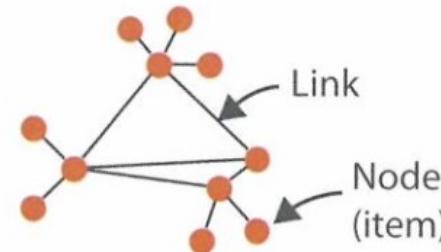
- expressiveness principle
    - match channel and data characteristics
  - effectiveness principle
    - encode most important attributes with highest ranked channels

# Basics. Terminology

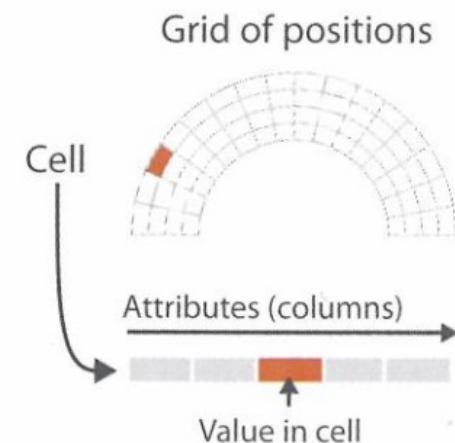
## → Tables



## → Networks



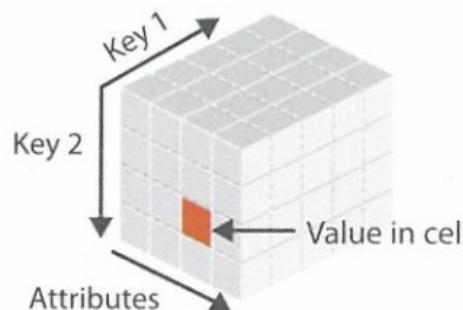
## → Fields (Continuous)



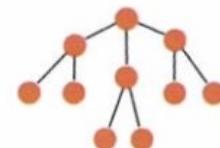
## → Geometry (Spatial)



## → Multidimensional Table

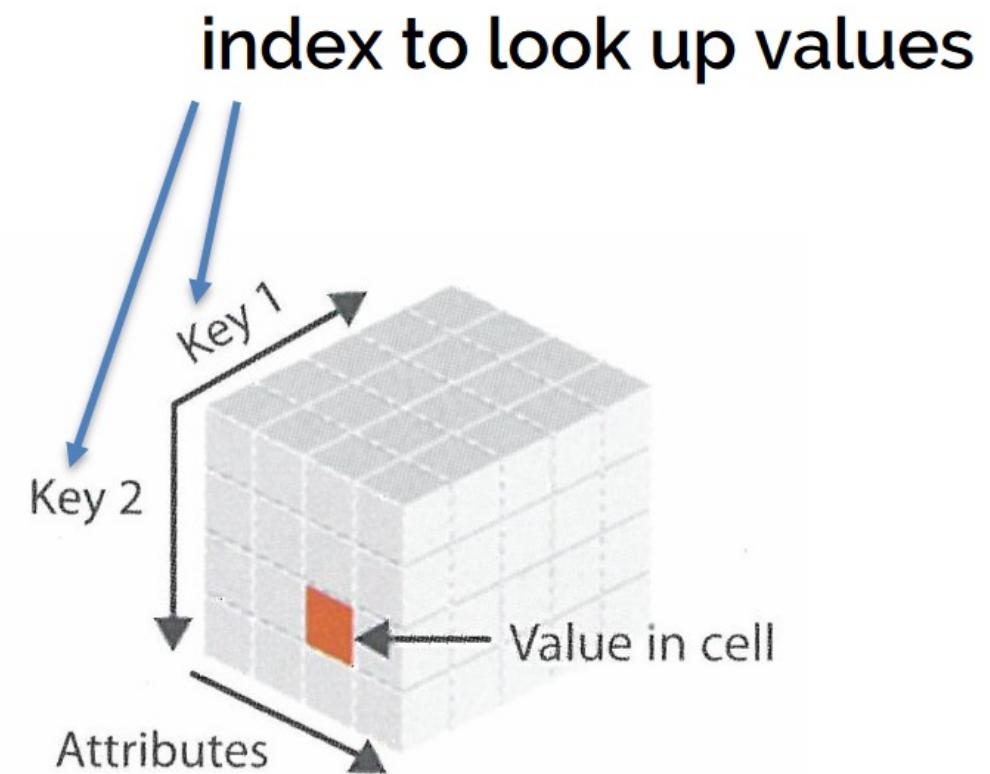
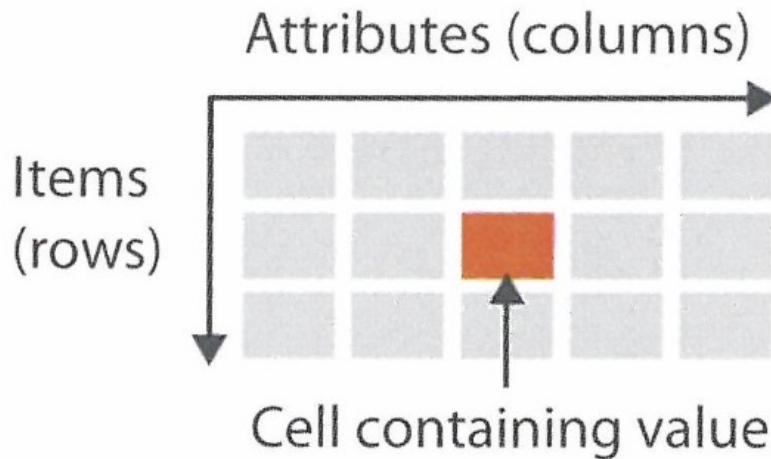


## → Trees



# Basics. Terminology

- What could be a key?
  - What data type is suitable for a key?



# Basics. Terminology

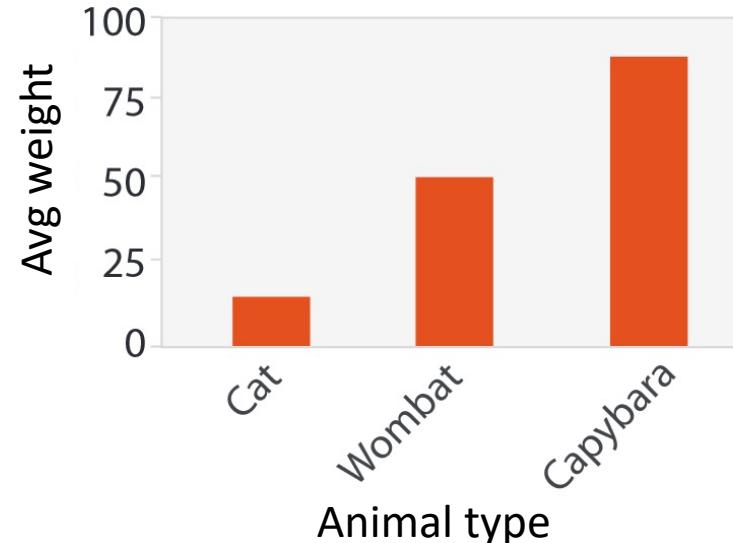
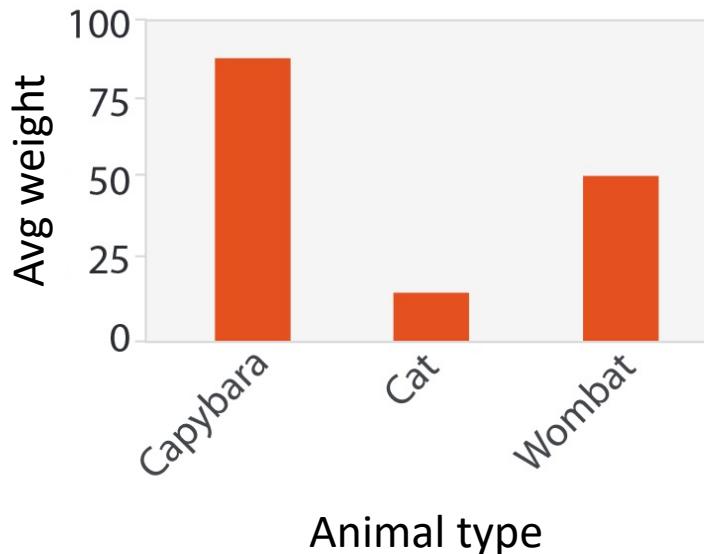
- Keys vs. Values
  - Key attributes are also sometimes called:
    - Independent attribute
    - Dimension
  - Value attributes are also sometimes called:
    - Dependent attribute
    - Measure
- Levels
  - Unique values for a categorical or ordered attribute

# Basic Data Visualization Techniques

Pere-Pau Vázquez

Dept. Computer Science – UPC

# Representations. Bar charts



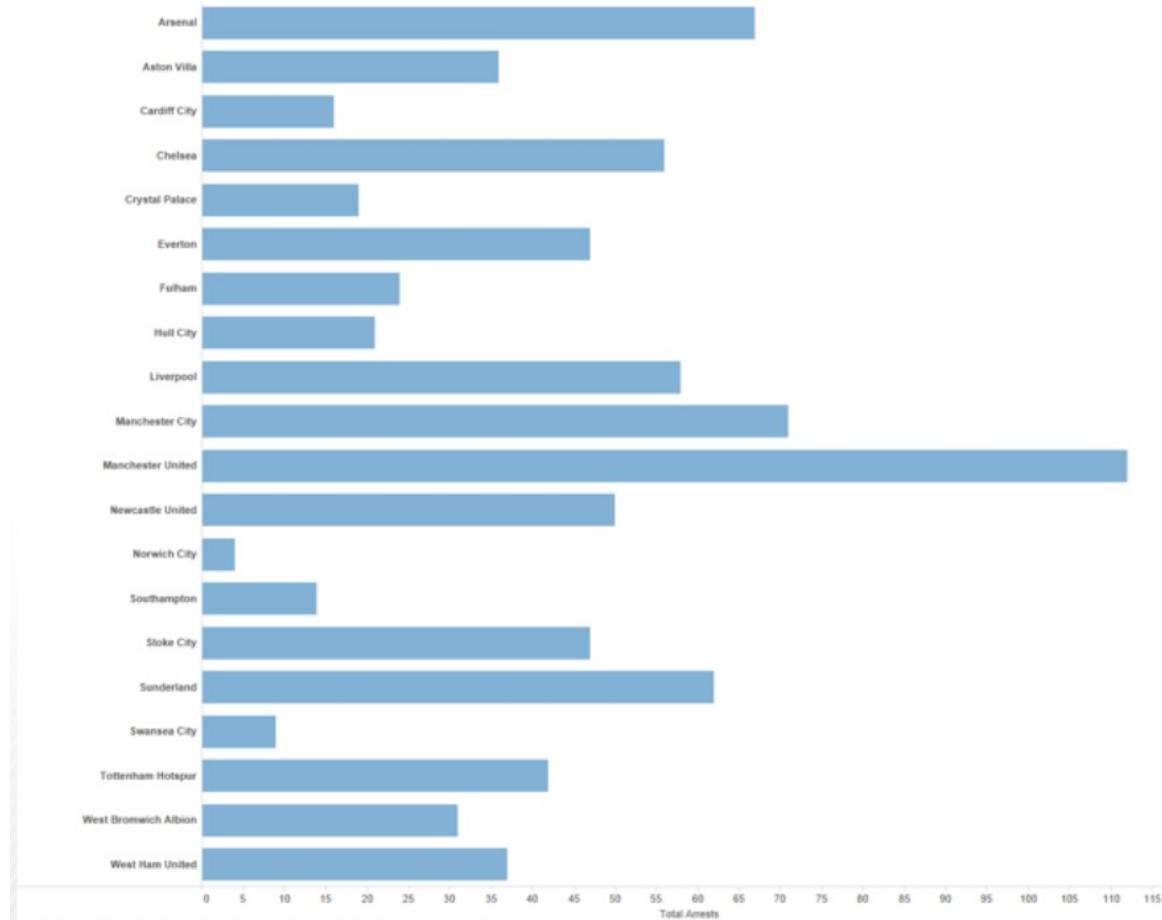
- One key, one value
  - Data:
    - 1 category attribute, 1 quantity attribute
  - Mark:
    - lines

# Representations. Bar charts

- One key, one value
  - Channels:
    - Length to express quantity
    - Spatial regions: one per mark
      - Separated horizontally, aligned vertically
      - Ordered by quantitative attribute
        - » By label (alphabetical), by length attribute (data-driven)
  - Task:
    - Compare, lookup values
  - Scalability
    - Dozens to hundreds of levels for key attribute

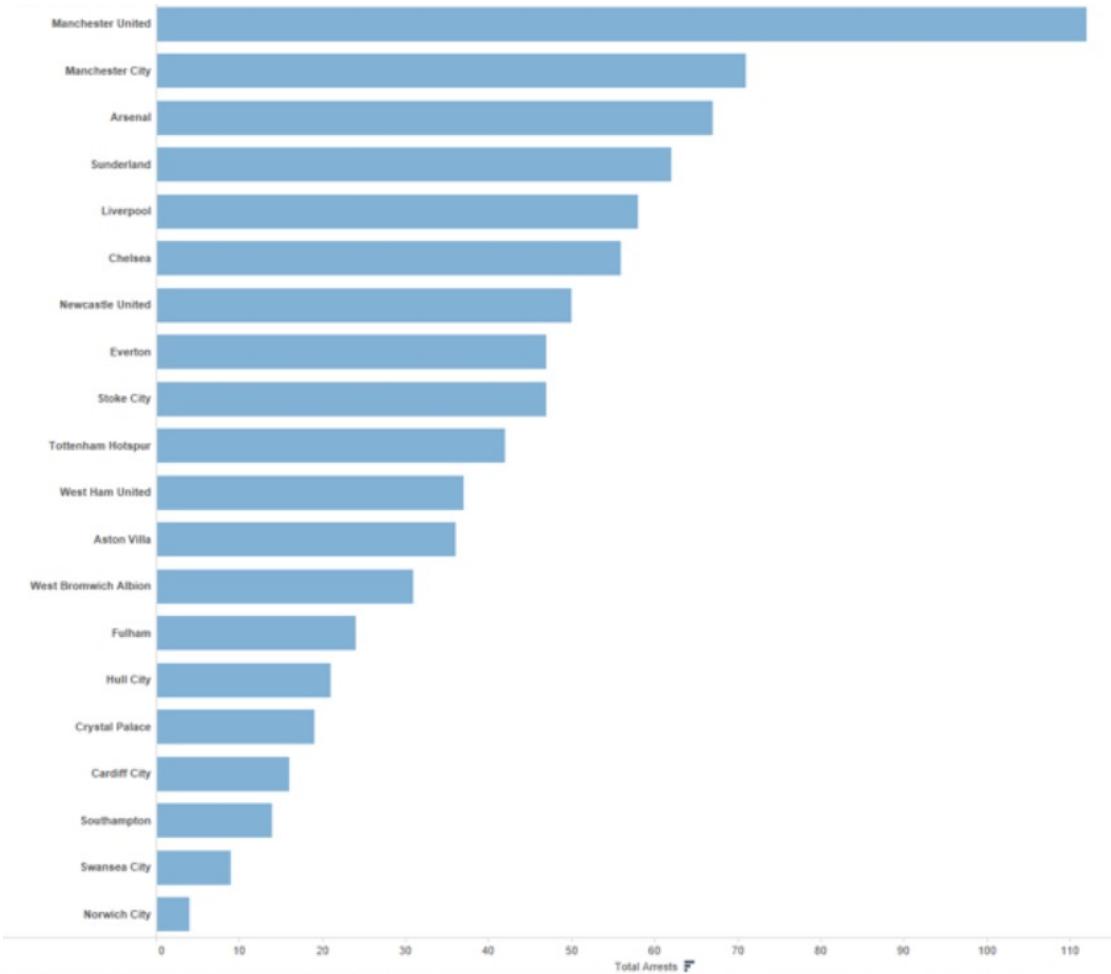
# Representations. Bar charts

- Separated and aligned,  
but not ordered
  - Hard to know rank

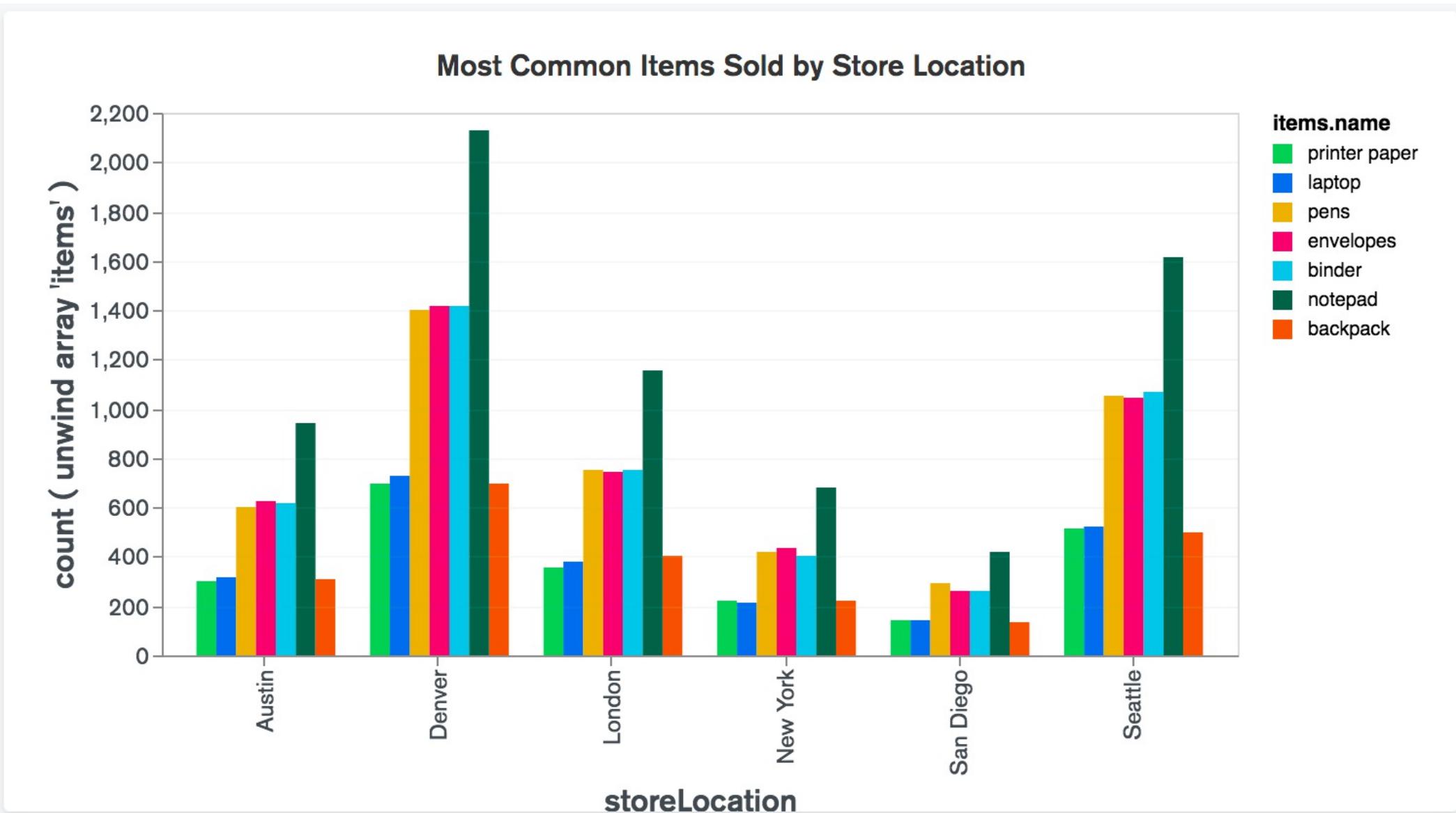


# Representations. Bar charts

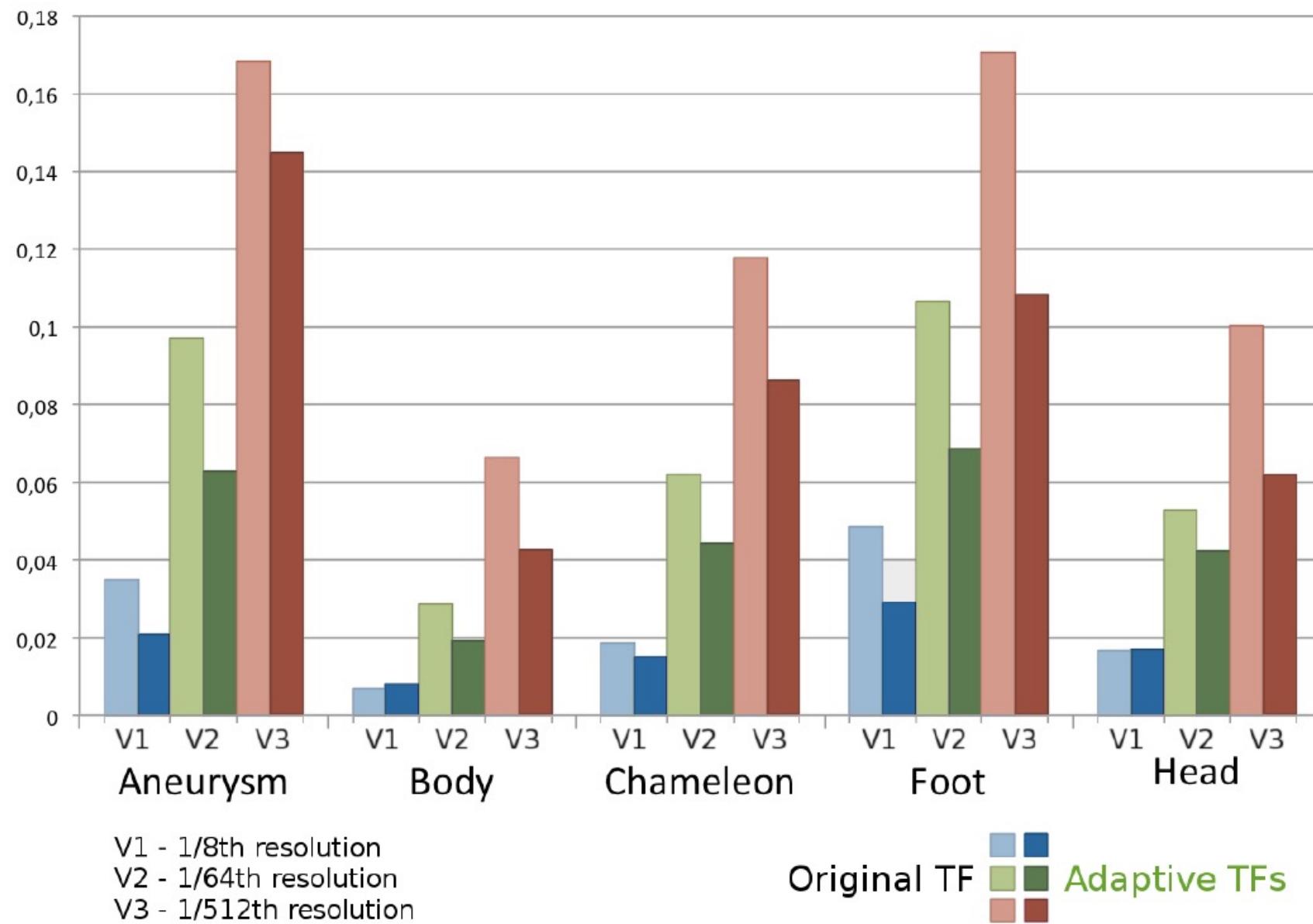
- Separated aligned, and ordered



# Representations. Paired/grouped bar charts



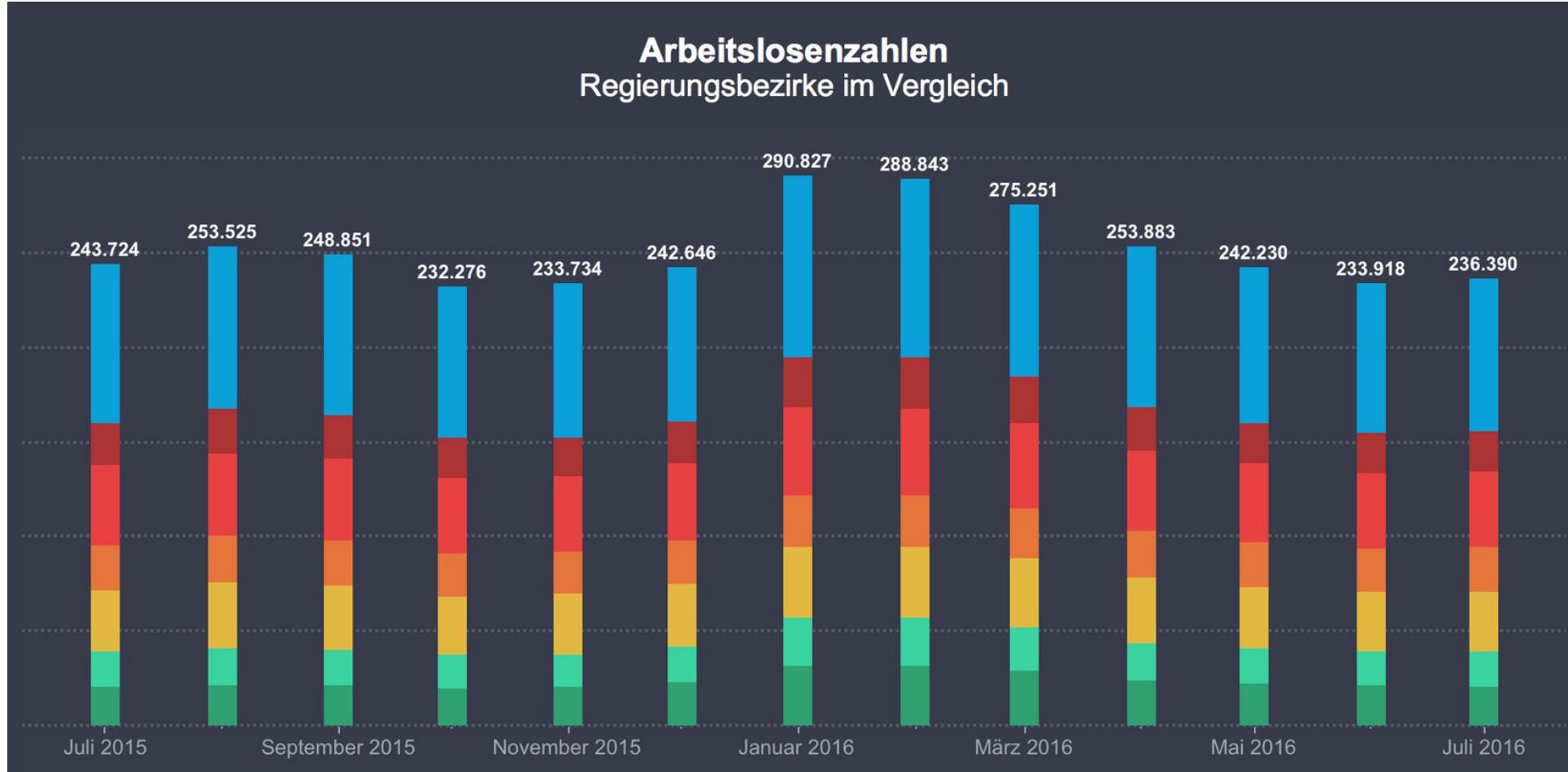
# Representations. Paired/grouped bar charts



# Representations. Paired/grouped bar charts

- One value, two keys
  - Channels:
    - Length to express quantity
    - Color hue for one key
    - Spatial regions for the other key
      - Separated horizontally, aligned vertically
      - Sometimes ordered by quant attribute, sometimes only on the keys
        - » By label (alphabetical)
  - Task:
    - Compare (within same key, among keys), lookup values
  - Scalability
    - Depends on the number values in the second key (< one dozen?), smaller than simple bar charts

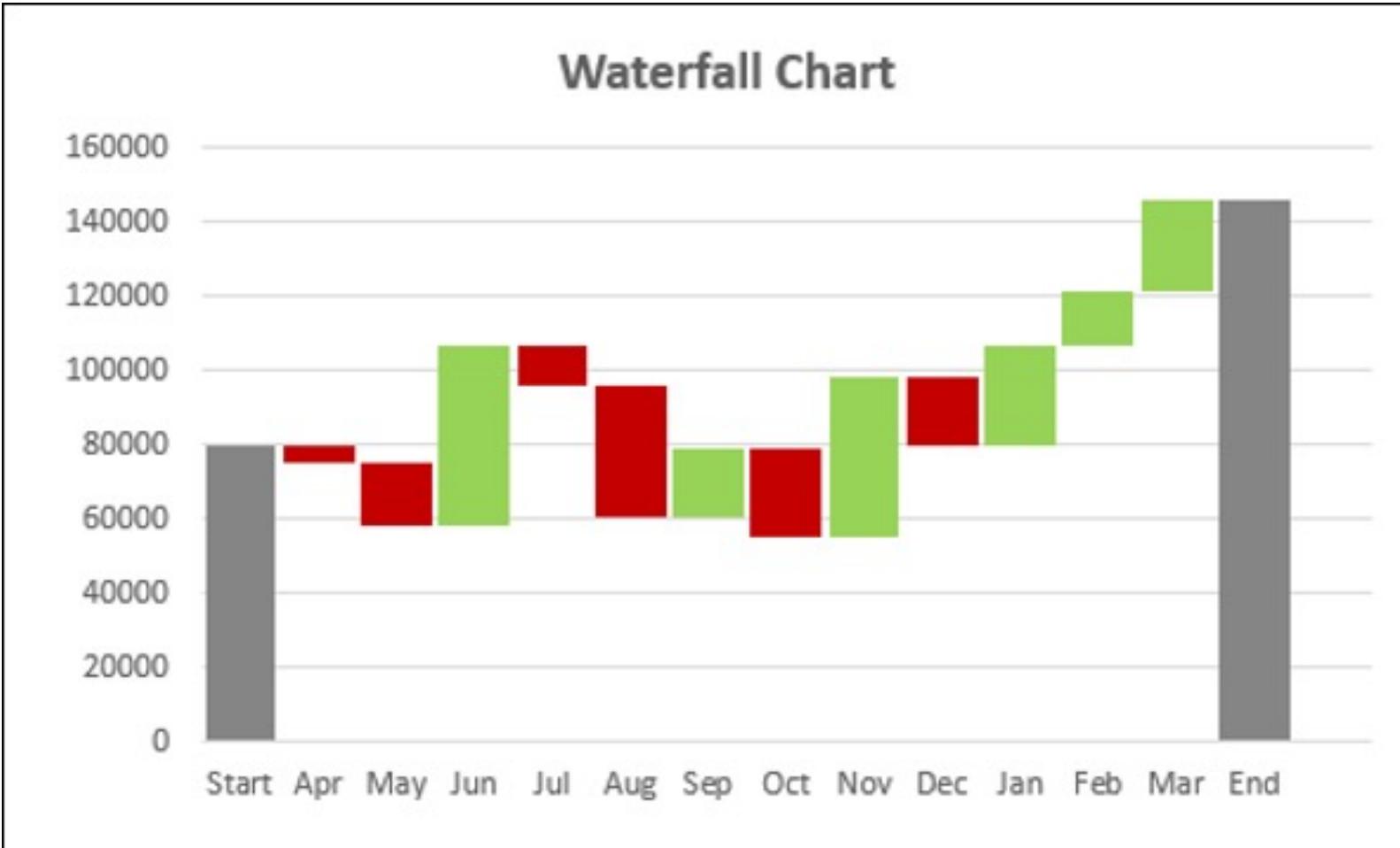
# Representations. Stacked bar chart



# Representations. Stacked bar chart

- One value, two keys
  - Data: 2 category attributes, 1 quantitative attribute
  - Mark: vertical stack of line marks
  - Channels
    - Length and color hue
    - Spatial regions: one per bar
      - Aligned: full concatenated bar, lowest bar component
      - Unaligned: other bar components
  - Task
    - Part-to-whole relationship, lookup values, find trends
  - Scalability
    - Several to one dozen levels for stacked attribute

# Representations. Waterfall chart



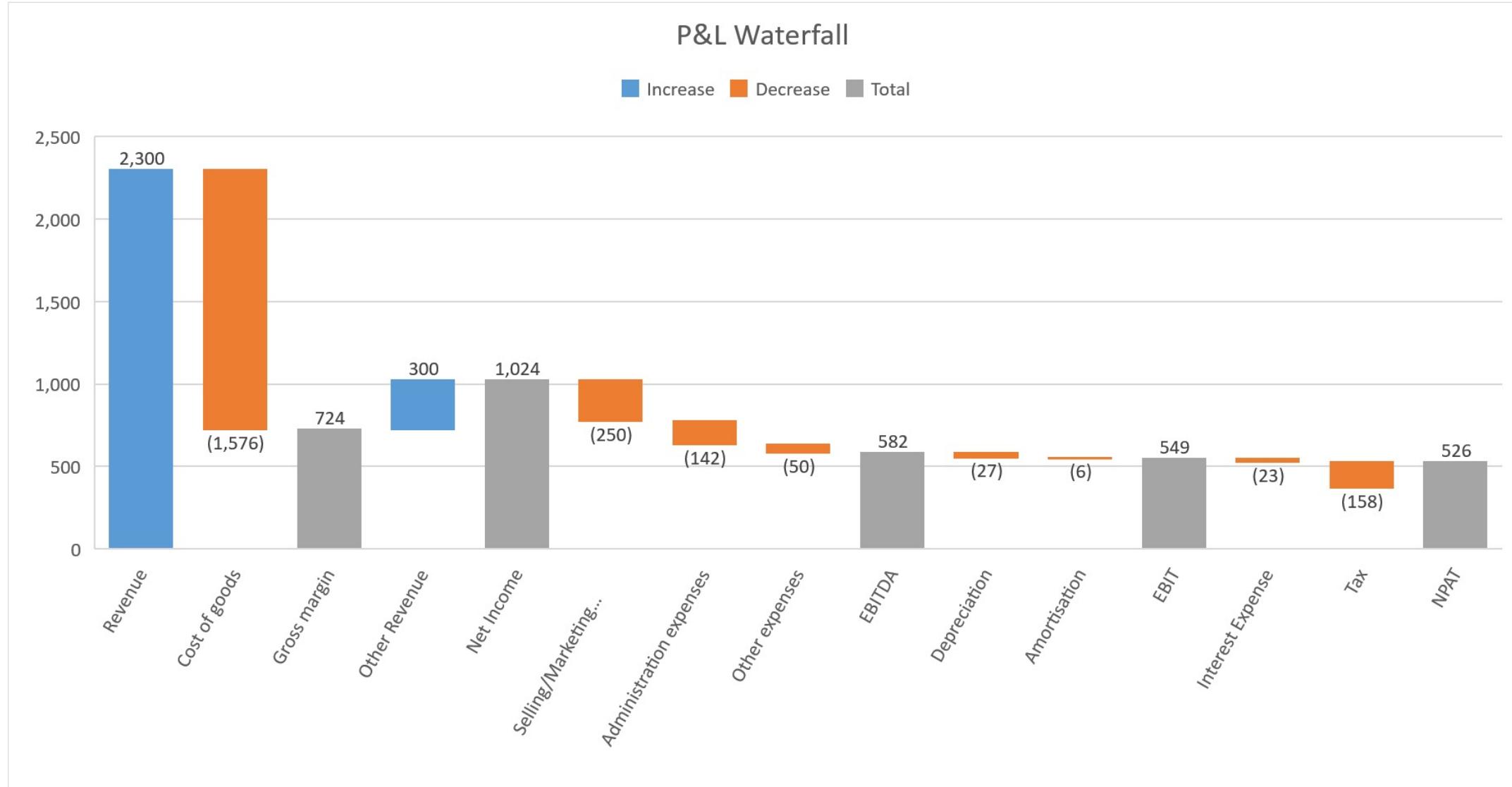
# Representations. Waterfall chart

- Useful to compare evolution
  - Suitable when positive and negative values are present
- Contrasting colors to highlight differences in data sets
- Can combine totals with evolution within the same chart
  - Ensure they are not confused

# Representations. Waterfall chart

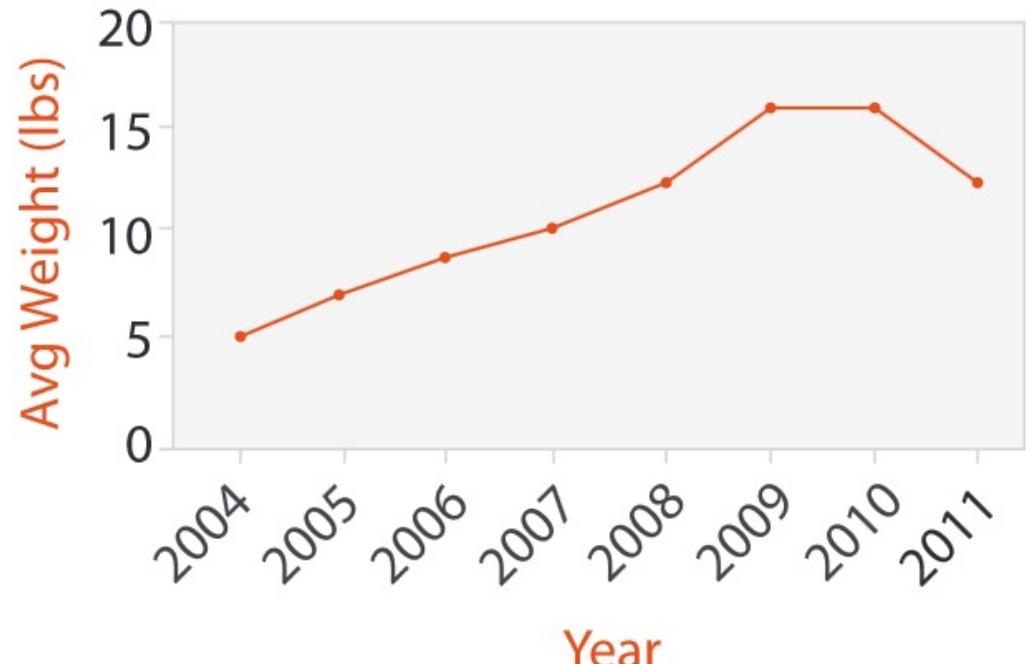
- One key (ordered), one value
  - Channels:
    - Length to express quantity
    - Hue to indicate grow/decrease/total
    - Spatial regions: one per mark
      - Separated horizontally, aligned vertically (last value)
      - Ordered by ordered key (typically temporal)
  - Task:
    - Understand evolution of a certain [initial] value, understand cumulative effect
  - Scalability
    - Dozens to hundreds of levels for key attribute

# Representations. Waterfall chart



# Representations. Line charts

- One key, one value
  - Data
    - 2 quant attribs
  - Mark: points
    - Line connection marks between them
  - Channels
    - Aligned lengths to express quant value
    - Separated and ordered by key attrib into horizontal regions
  - Task:
    - Find trend
      - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next



# Representations. Bar vs line charts

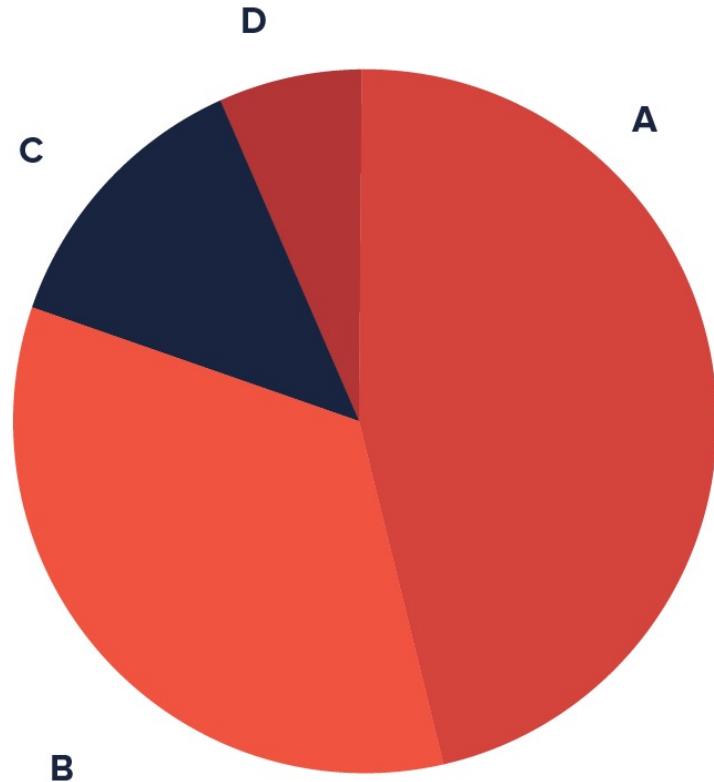
- What representation is better?

# Representations. Bar vs line charts

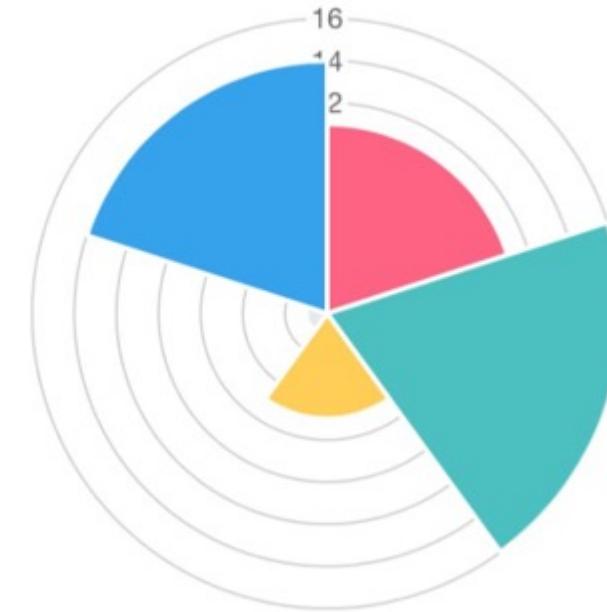
- Depends on type of key attribute
  - Bar charts if categorical
  - Line charts if ordered
- Do not use line charts for categorical key attributes
  - Violates expressiveness principle
    - Implication of trend so strong that it overrides semantics!, e.g. “The more male a person is, the taller he/she is”

# Representations. Pie chart, polar area charts

- Pie & Polar [area] charts



Red Green Yellow Grey Blue



# Representations. Pie chart, polar area charts

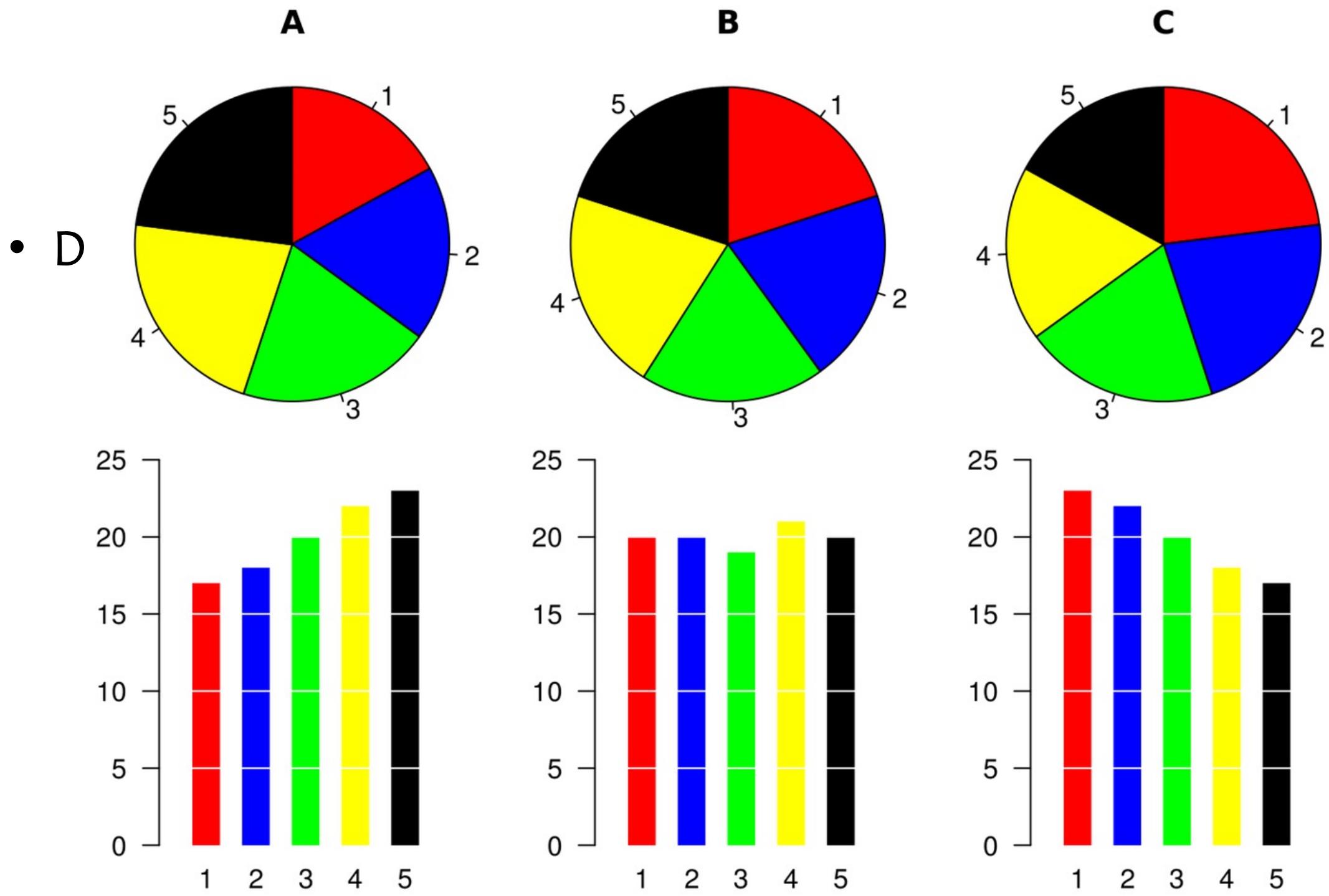
- Pie chart
  - Area marks with **angle** channel (*angle* encodes quantity)
  - Accuracy: angle/area much less accurate than line length
- Polar area chart
  - Area marks with **length** channel (*length* encodes quantity)
  - More direct analog to bar charts
- Data
  - 1 key attribute, 1 quantitative value attribute
- Task
  - Part-to-whole judgements

# Representations. Pie chart, polar area charts

- Discussion. Pie charts:
  - Main task: Part-to-whole judgments
    - If comparing numbers, a bar chart may be more suitable
  - Small number of slices (less than 10)
  - Values have to differ
  - Sort the values (seems to be better)
  - Start at 12 o'clock (seems to be better)

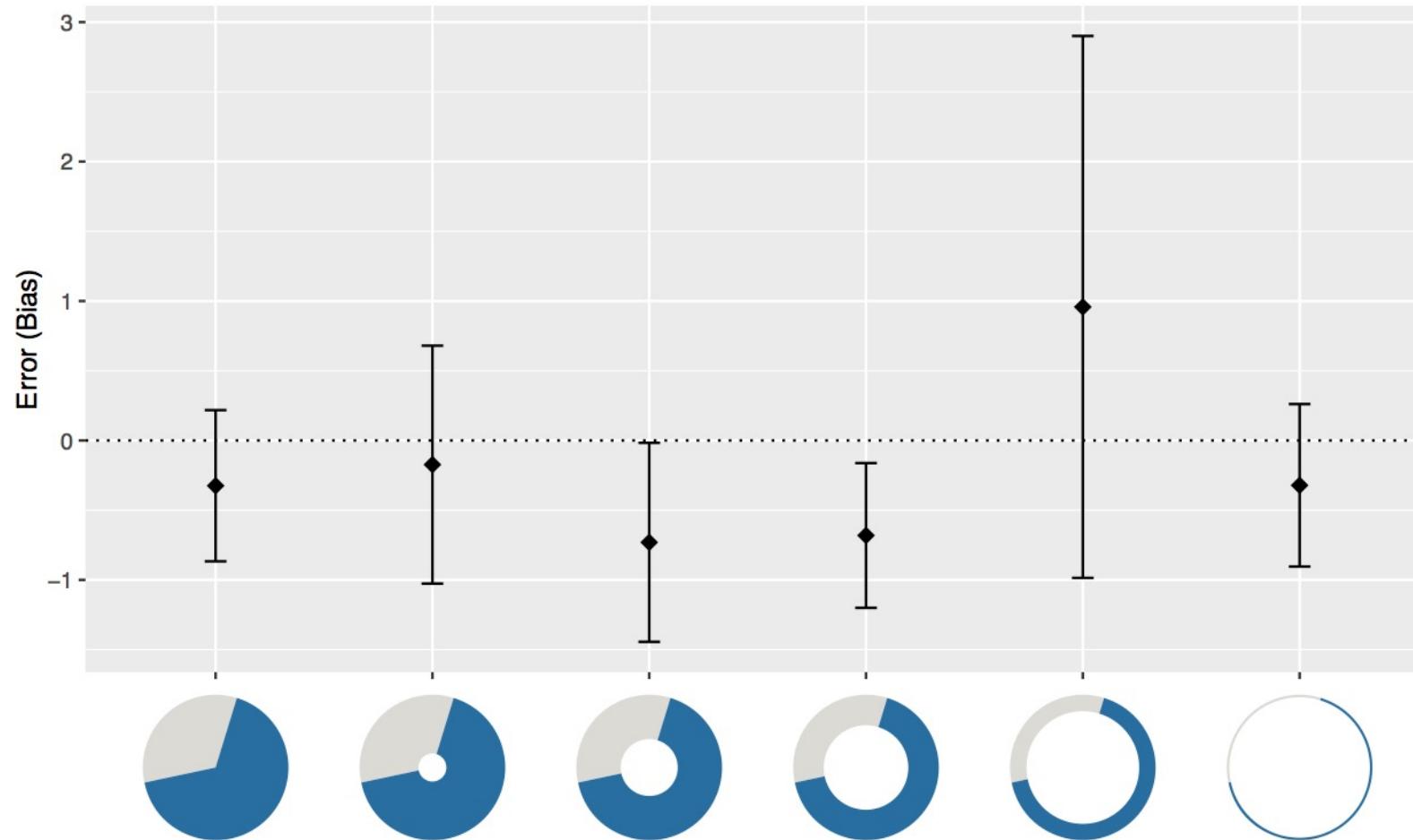
# Representations. Donut chart

- Donut charts:
  - Pie charts with a hole
  - Quantity also encoded in angle



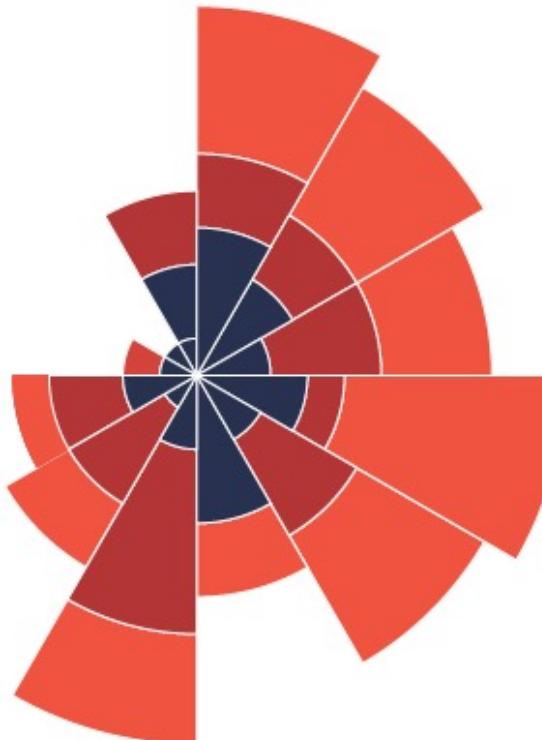
# Representations. Pie chart, polar area charts

- Donut charts seem to be equally good (cf. [Kosara 2016])



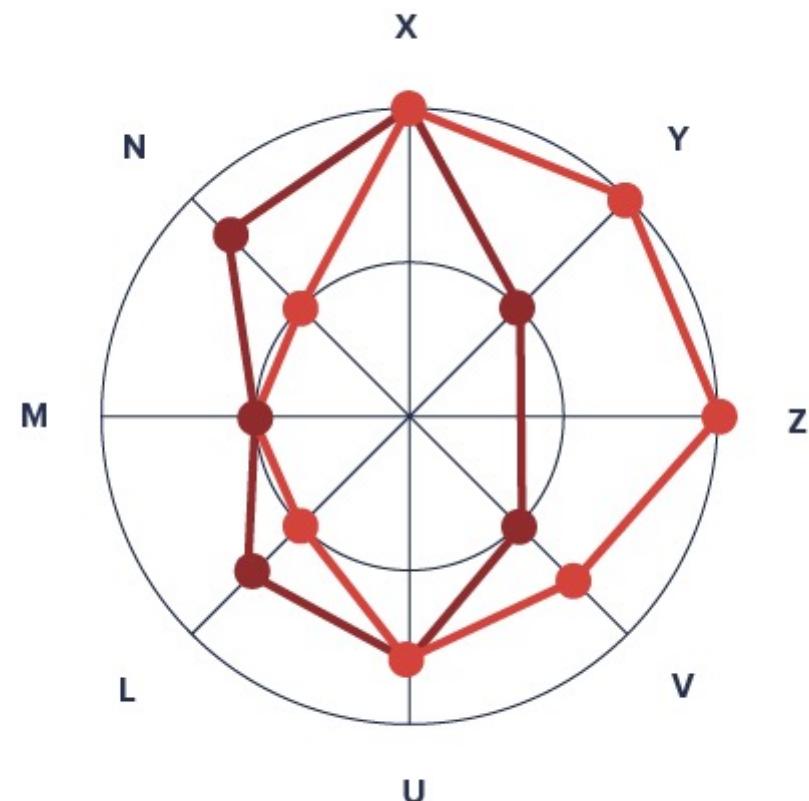
# Representations. Pie chart, polar area charts

- Can have all sorts of stacked versions:
  - Same problems (or worse) than with pie charts



# Representations. Radar chart

- Analogous to paired/grouped column charts
  - Multiple dimensions
  - Space efficient



# Representations. Radar chart

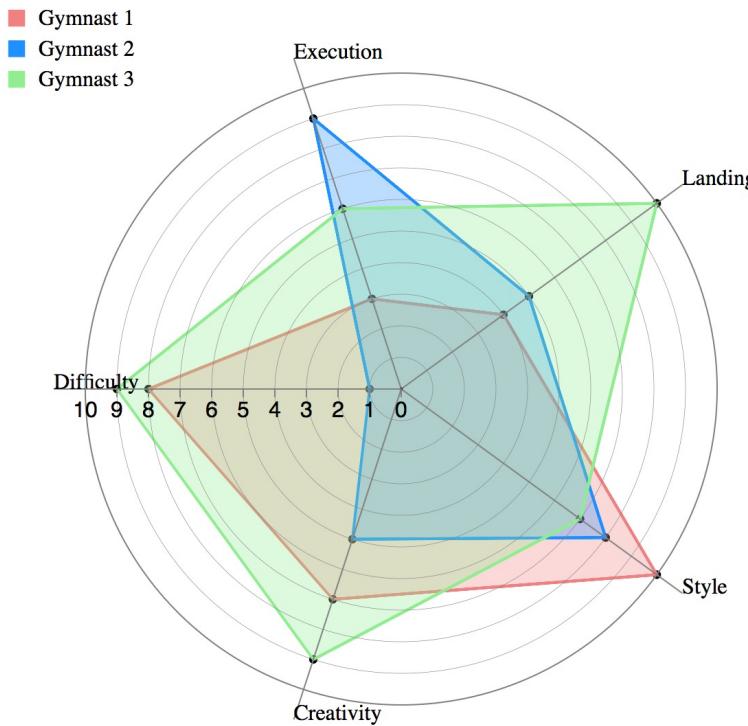
- Radar chart
  - Multiple dimensions
- One value, two keys
  - Data: 2 category attributes, 1 quantitative attribute
  - Mark: distance from the center of line marks
- Task
  - Comparison of multiple variables

# Representations. Radar chart

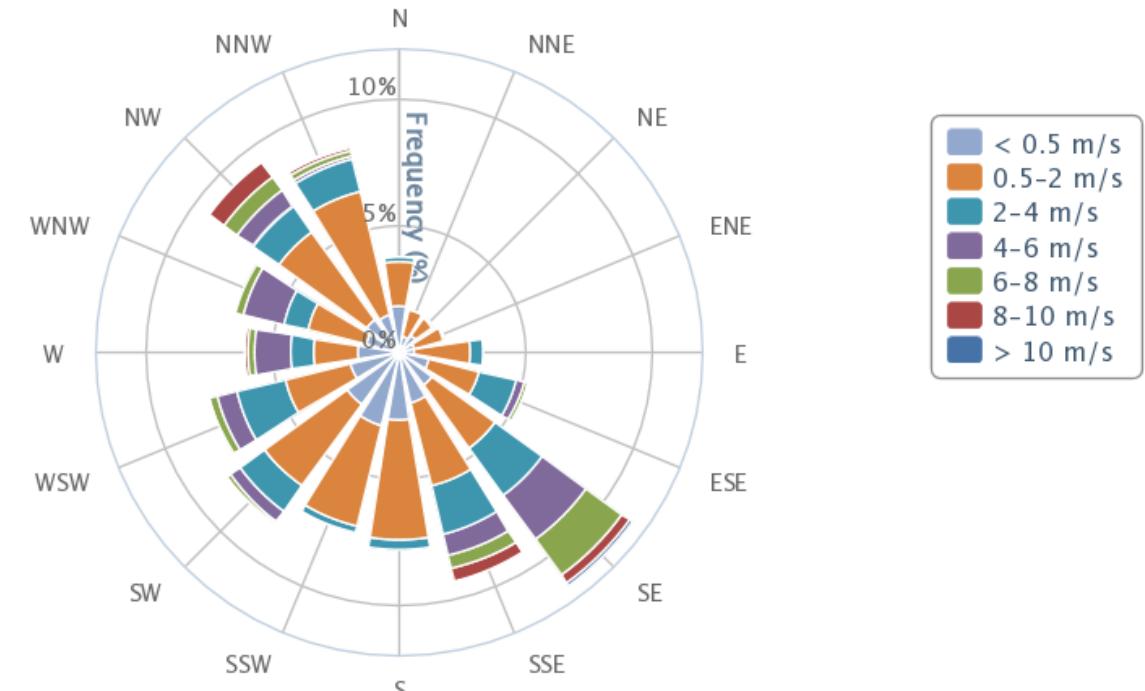
- Other names: spider chart, star charts
- Many variations:
  - Areas, stacks (like polar area charts... )

# Representations. Radar chart

Gymnast Scoring Radar Chart



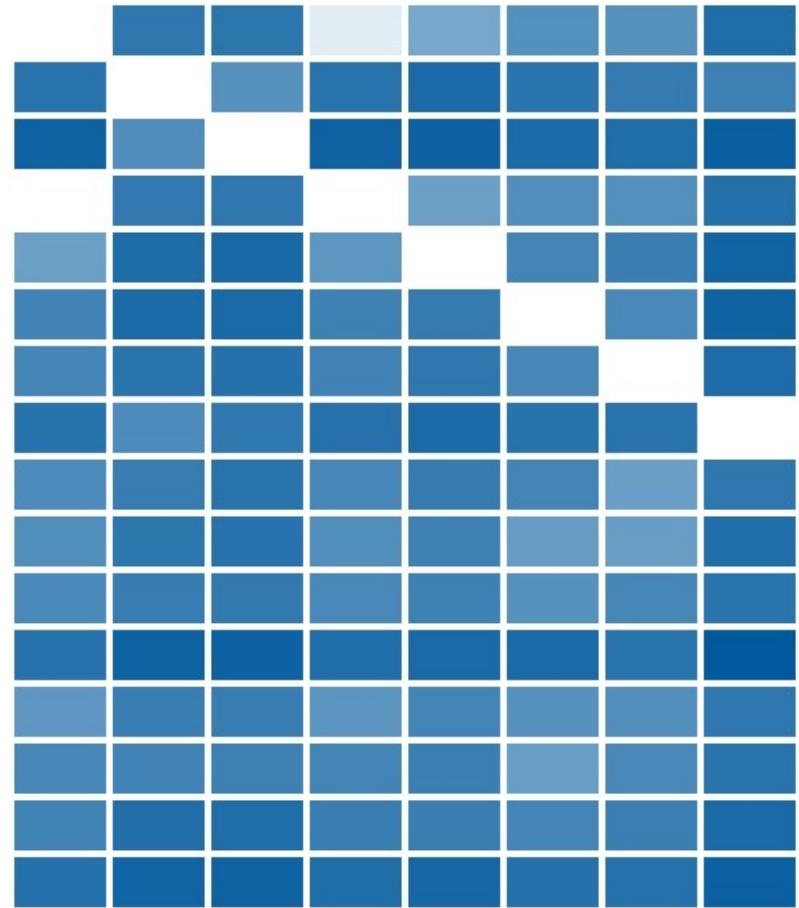
Wind rose for South Shore Met Station, Oregon  
Source: or.water.usgs.gov



Highcharts.com

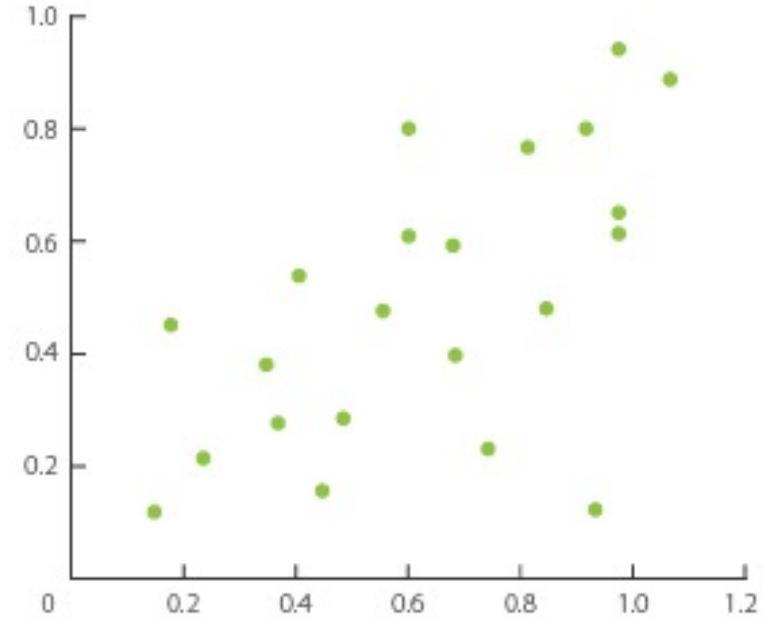
# Representations. Heat maps

- Two keys, one value
  - data
    - 2 categorical attributes (article1, article2)
    - 1 quant attribute (difference)
  - Marks: area
    - Separate and align in 2D matrix
      - Indexed by 2 categorical attributes
  - Channels
    - Color by quantitative attribute (ordered colormap)
  - Task
    - Find clusters, outliers
  - Scalability
    - 1M items, 100s of category levels, ~10 quant attribute levels



# Representations. Scatterplots

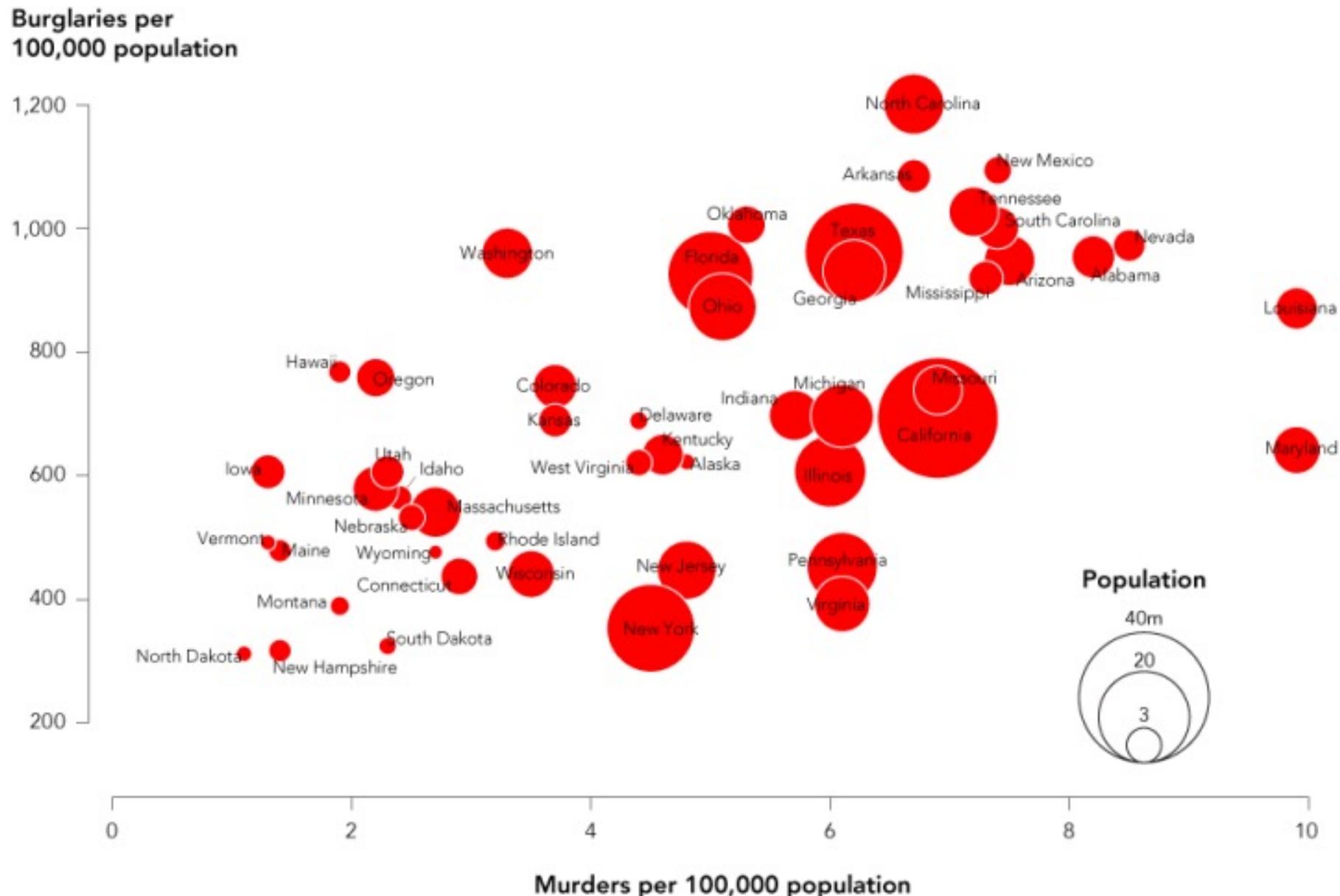
- Expresses values
  - Quantitative attributes
- No keys, only values
  - Data
    - 2 quant attributes
  - Mark: points
  - Channels
    - Horizontal + vertical position
  - Tasks
    - Find trends, outliers, distribution, correlation, clusters
  - Scalability
    - Hundreds of items



# Representations. Bubble chart

- Scatterplot + size of bubble (circle)
  - Adds dimension
  - Areas more difficult to compare
    - Scale bubbles according to area, not diameter
    - Use only circular shapes

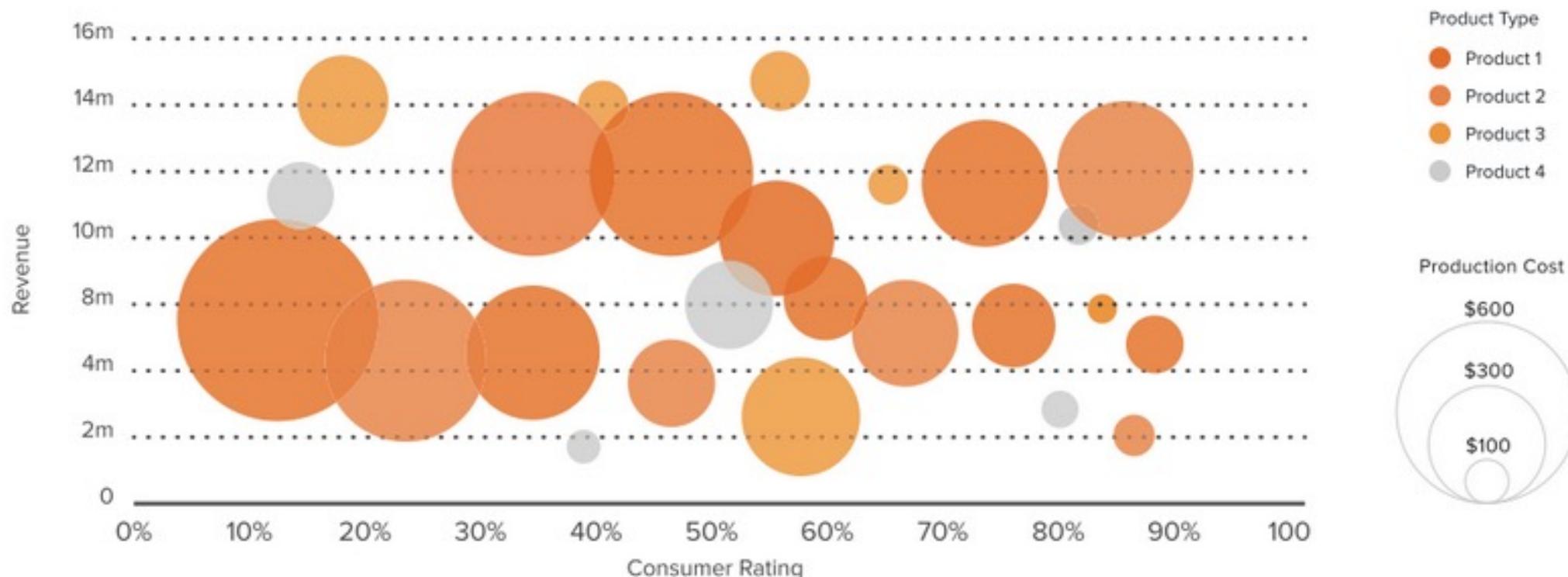
# Representations. Bubble chart



# Representations. Bubble chart

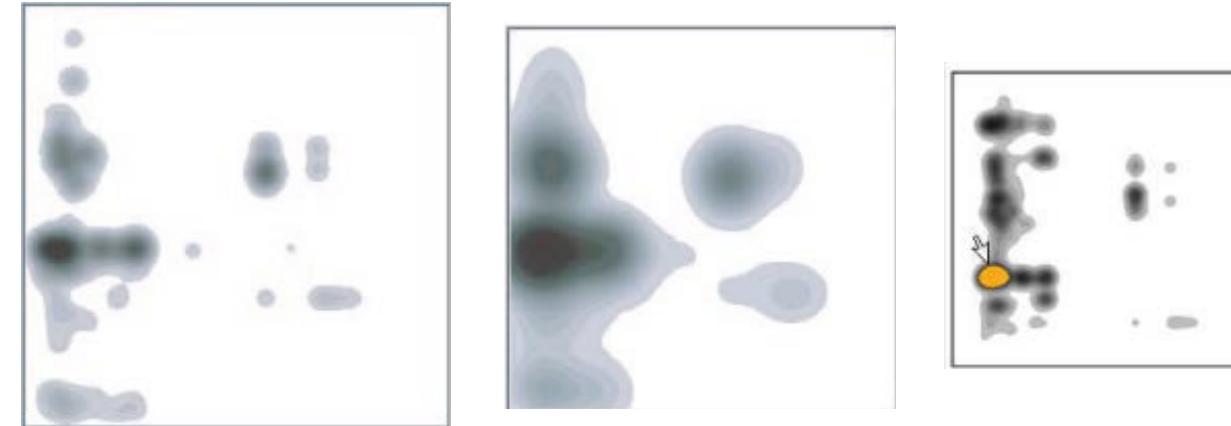
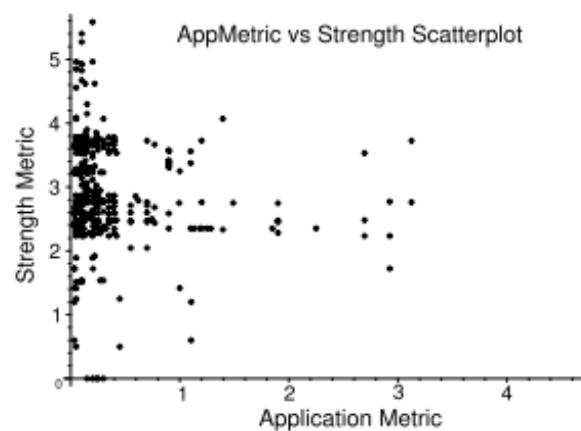
- Can extend to more attributes

## REVENUE VS. RATING



# Representations. Multiscale scatterplots

- Smooth scatterplot by convolving with Gaussian at different scales.
  - Facilitates finding patterns at different scales
  - Patches (groups) easily determined with image segmentation

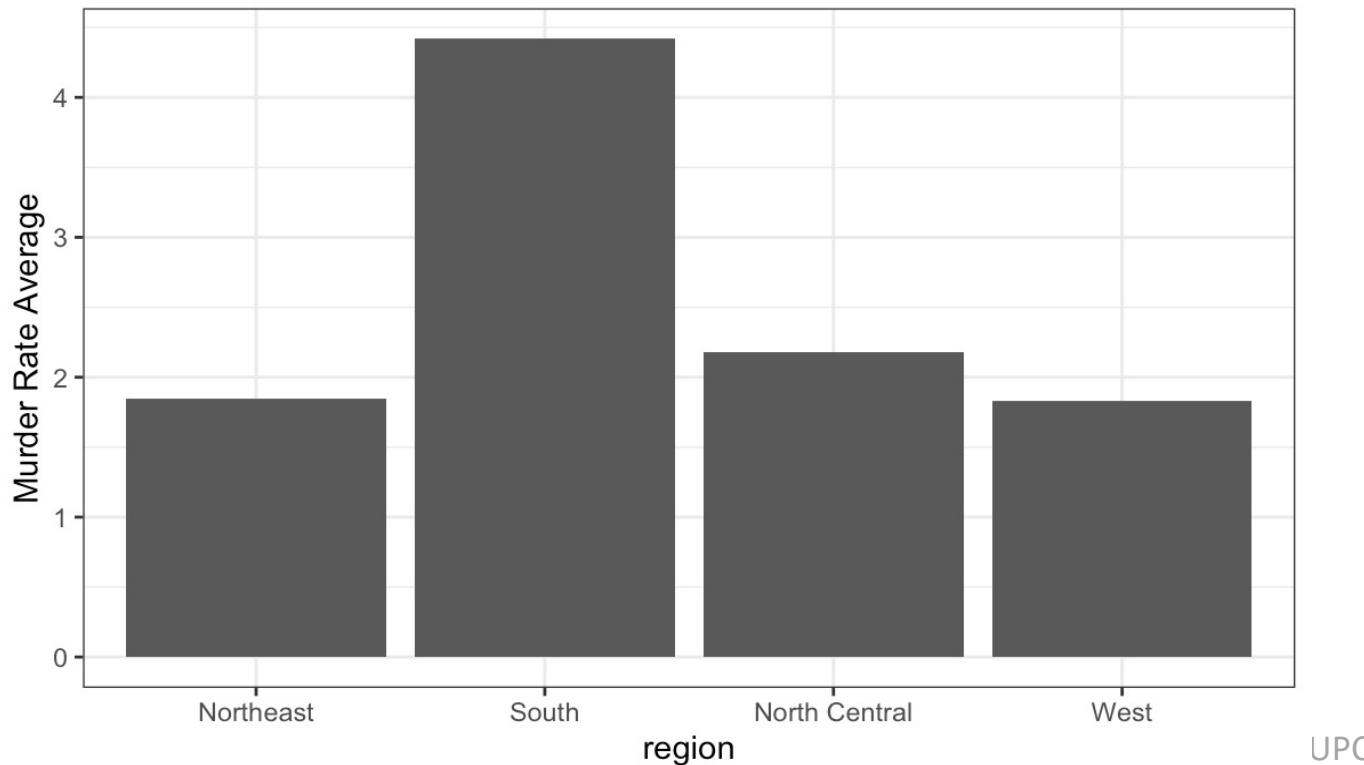


# Representations. Stem-and-leaf

- Alternative to histogram
    - Bins numbers according to the first significant digit
      - Uses the data to *paint* the representation

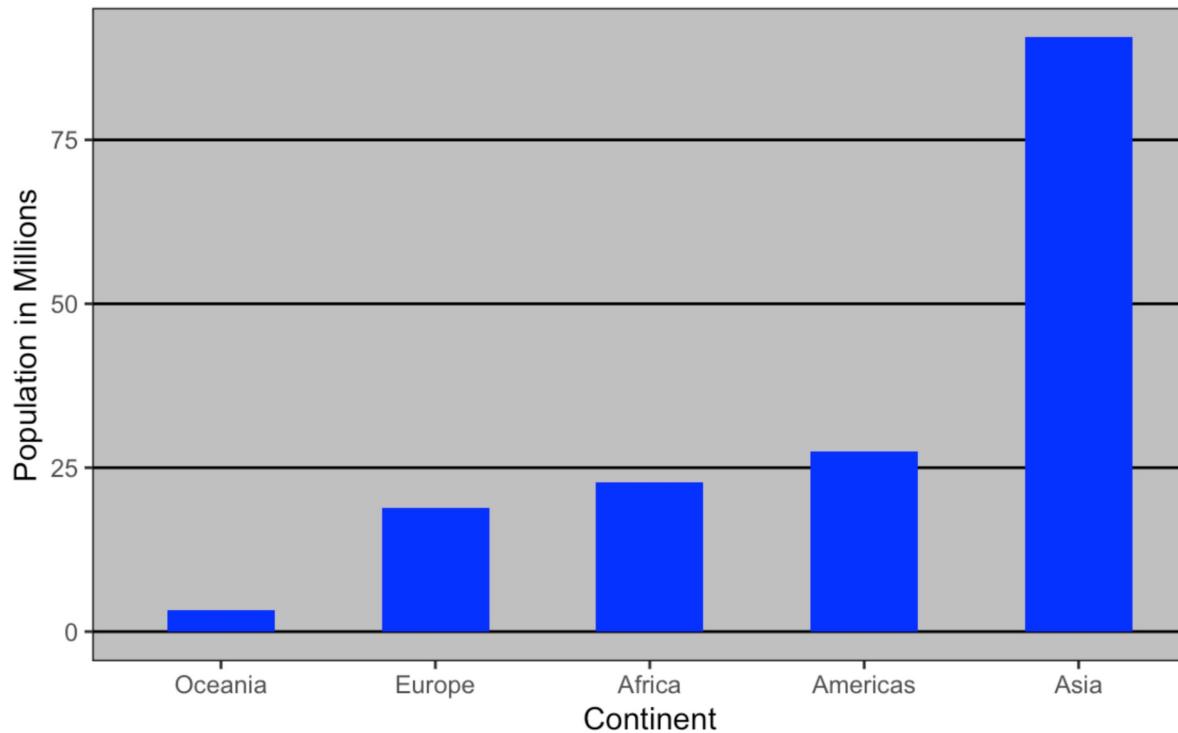
# Representations. Exercise

- Say we are interested in comparing gun homicide rates across regions of the US. We see this plot:
  - What region would you consider safer? Why? Discuss its design.



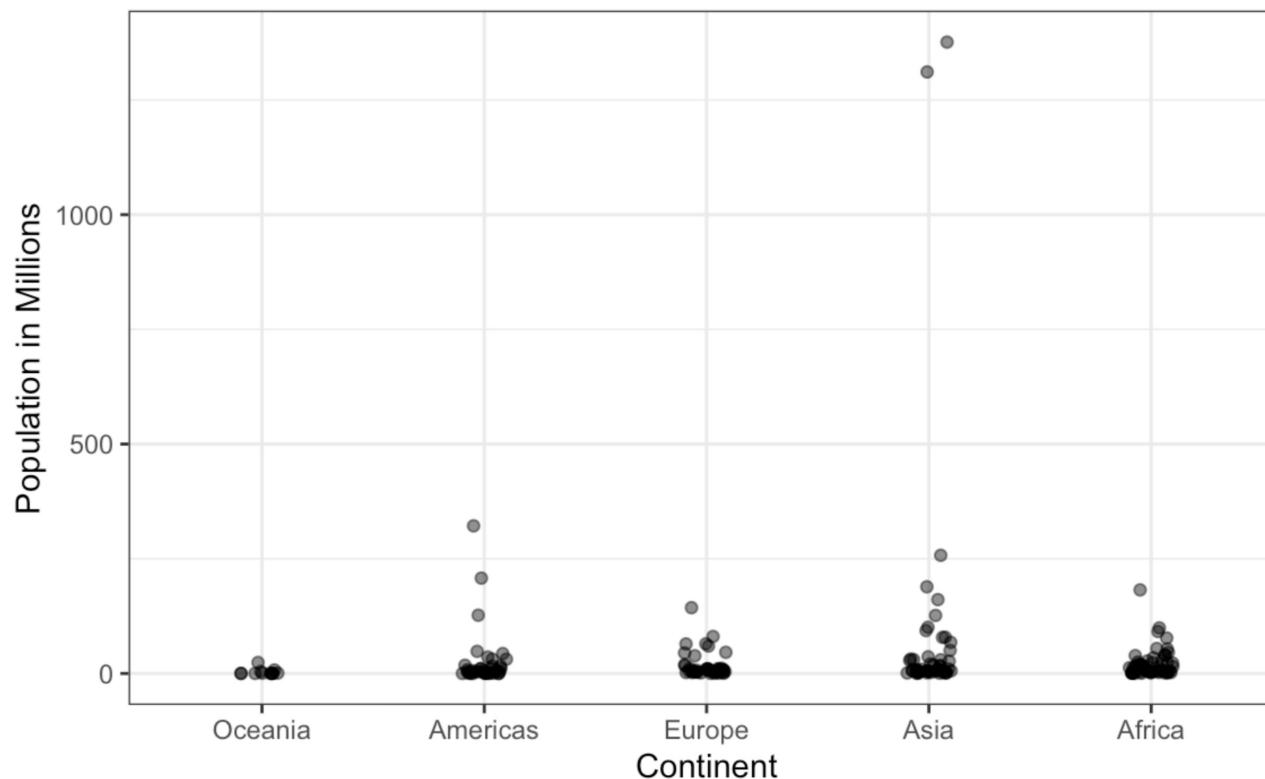
# Representations. Exercise

- Consider the following chart that shows the population of each continent in 2015. Do you think it is faithfully conveying the data? Why?



# Representations. Exercise

- What if we represent the data in the following way? Do you think it is effective at showing the data? Why?



# Representations. Exercise

Immigrants as a percentage of population in 2016, by period of immigration

● Before 1981 ● 1981-90 ● 1991-2000 ● 2001-10 ● 2011-16

30%

20

10

0

Thunder Bay census  
metropolitan area

Ontario

Canada

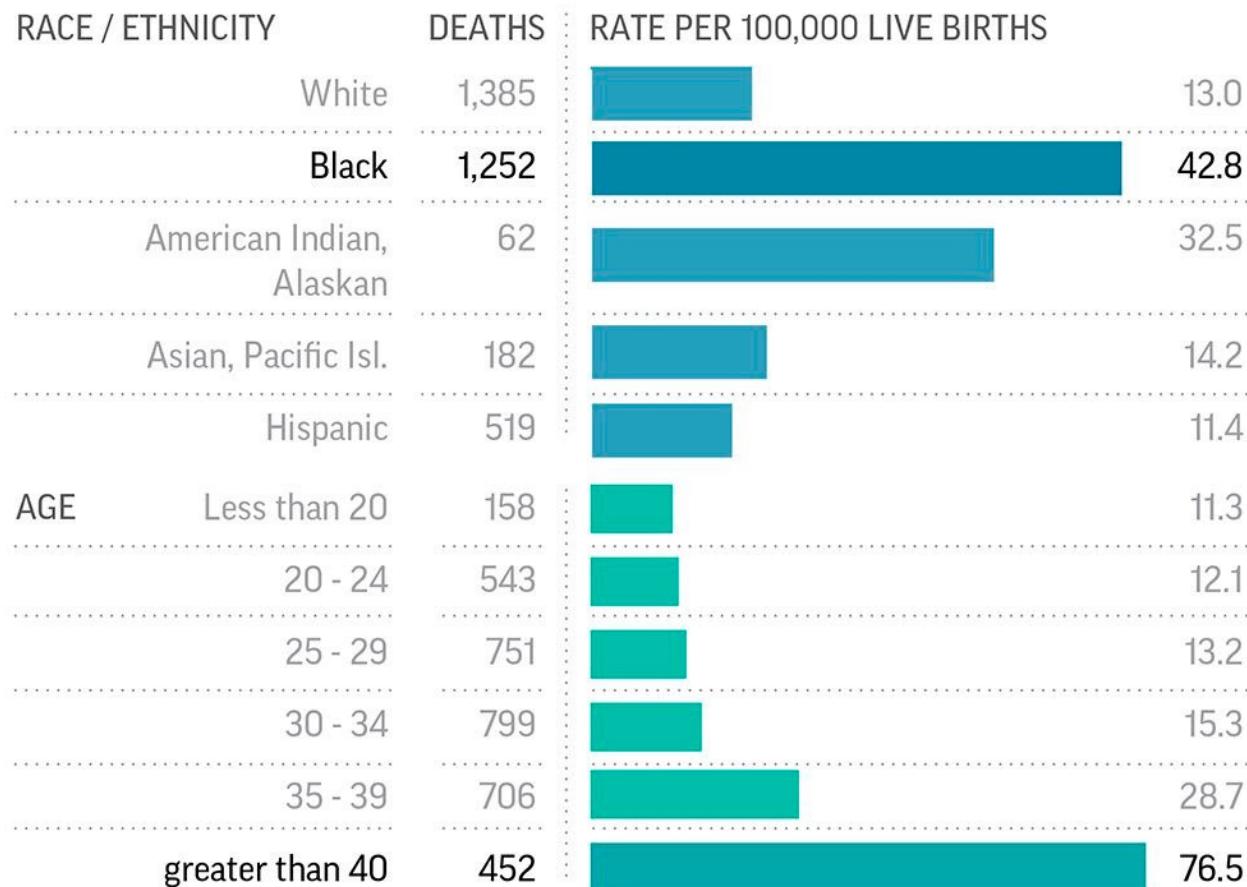
THE GLOBE AND MAIL, SOURCE: STATSCAN

DATA SHARE

# Representations. Exercise

Pregnancy deaths rare but higher in some groups

A new federal report finds that pregnancy-related deaths are rising in the U.S., especially among black women.



In 1 SOURCE: Centers for Disease Control and Prevention, 2011-2015 data

# Basic Visualization Techniques

Pere-Pau Vázquez

Dept. Computer Science – UPC

# Advanced Visualization Techniques

Pere-Pau Vázquez

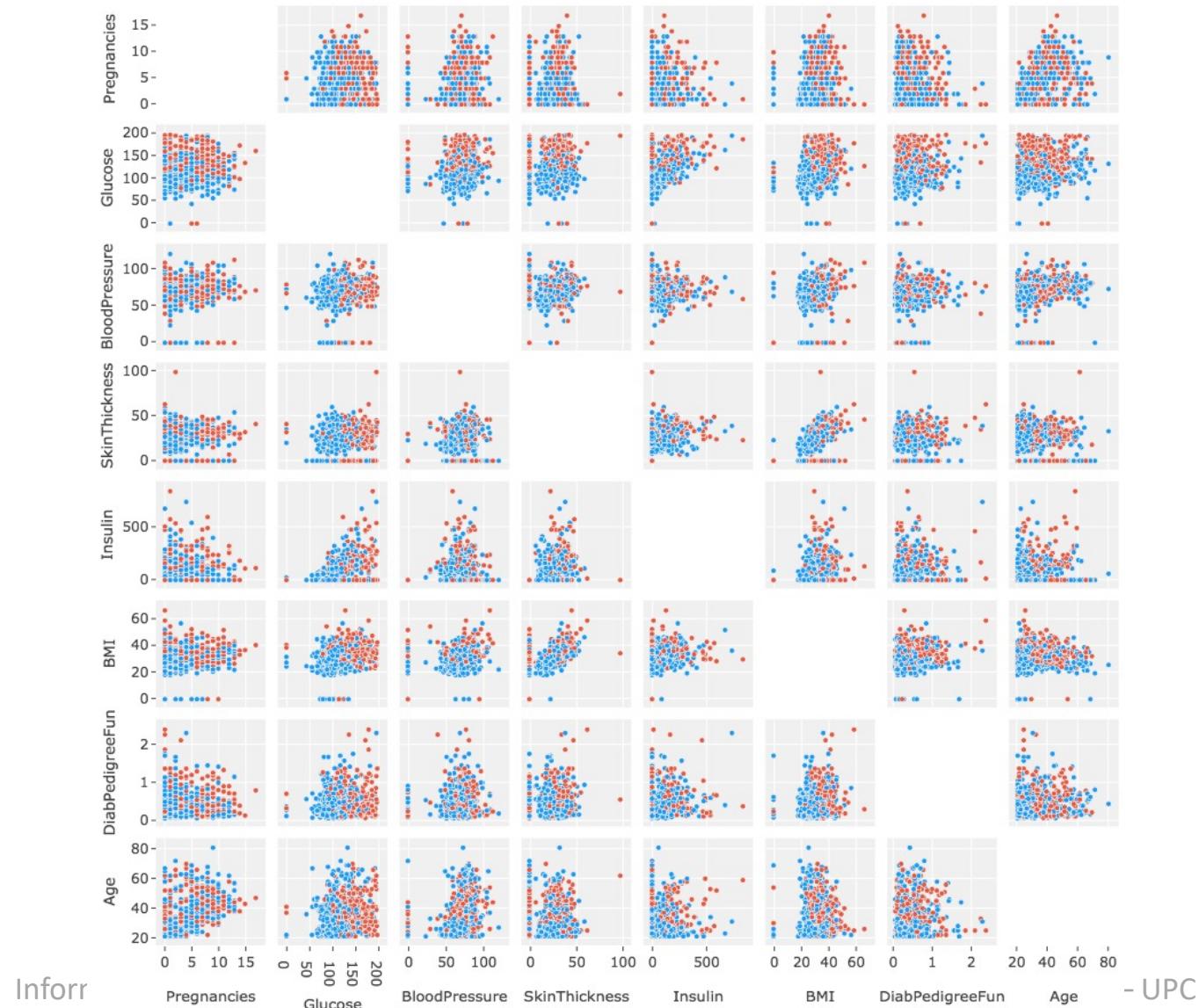
ViRVIG Group

Dept. Computer Science – UPC

# Outline

- Advanced data representations
- Multi-functioning elements
- Further reading

# Representations. Scatterplot matrices (SPLOM)



# Representations. Scatterplot matrices (SPLOM) + interaction

- Matrix showing all relations between variables
  - Selection / brushing, drilling down, linked brushing
- Tasks:
  - Find correlations between pairs of variables

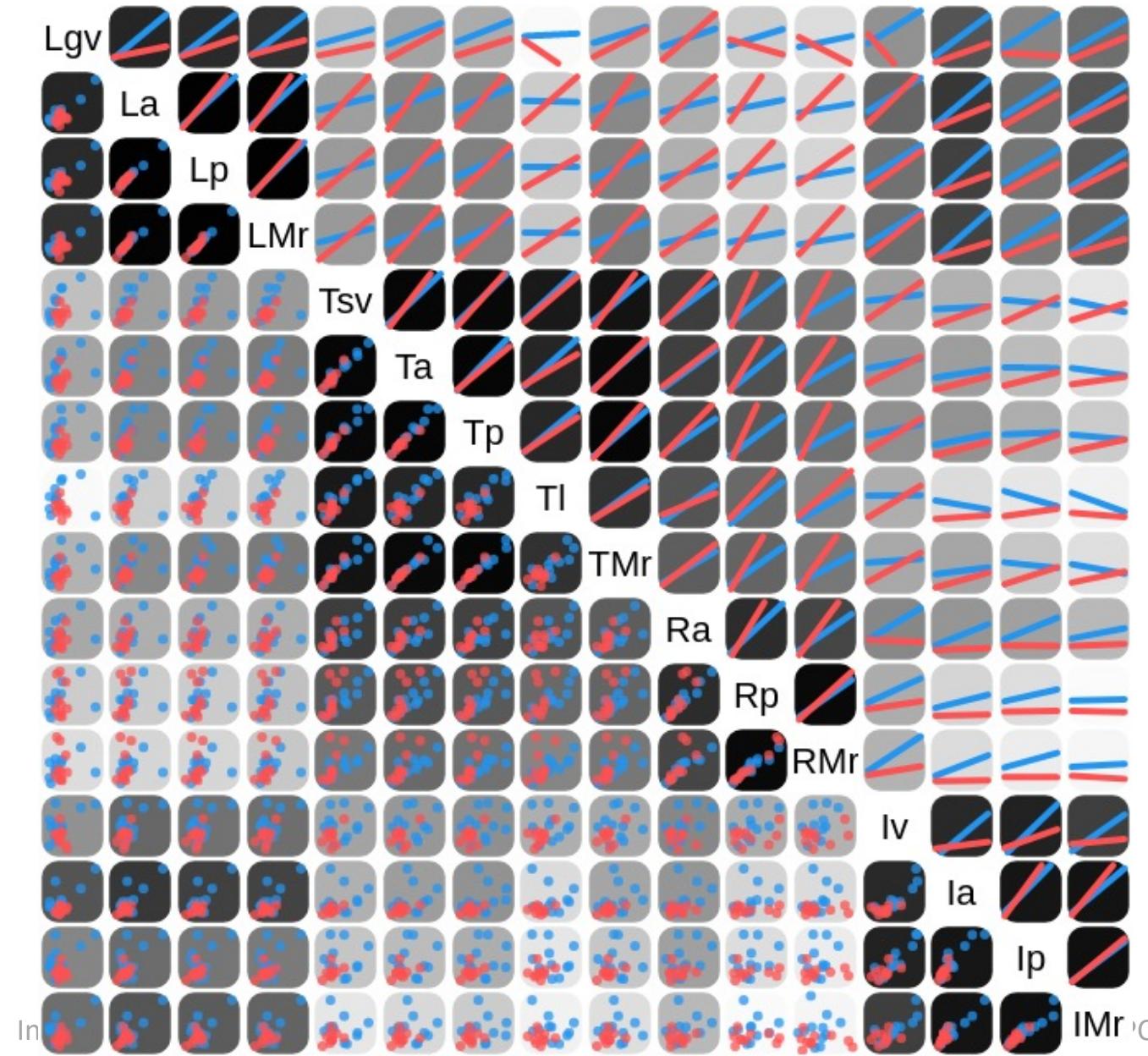
# Representations. Scatterplot matrices (SPLOM) + interaction

- Scatterplot matrix (SPLOM)
  - Rectilinear axes, point mark
  - All possible pairs of axes
  - Scalability
    - One dozen attributes
    - Dozens to hundreds of items

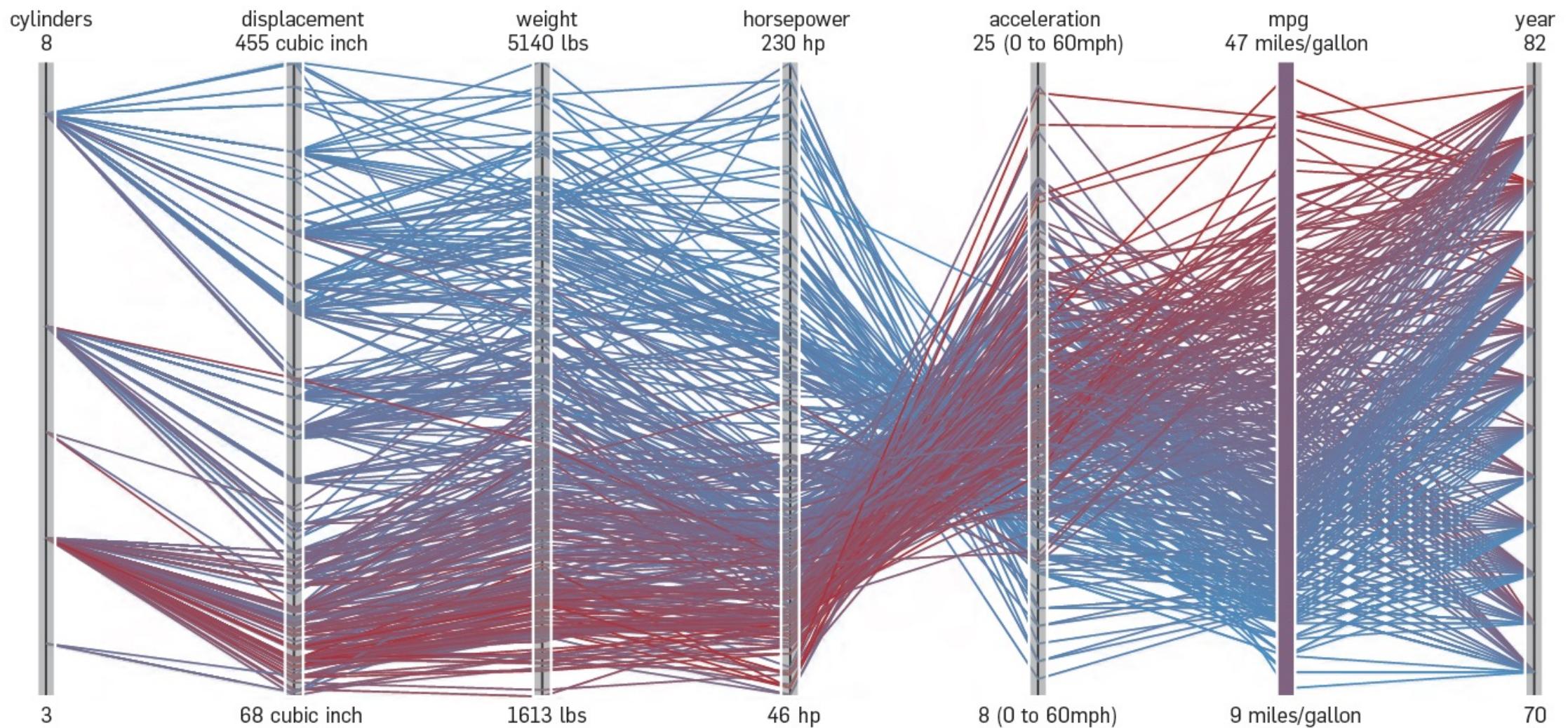
# Representations. Scatterplot matrices (SPLOM)

- Discussion:
  - Space
  - Repeated charts
  - Diagonal

# Representations. Scatterplot matrices (SPLOM)

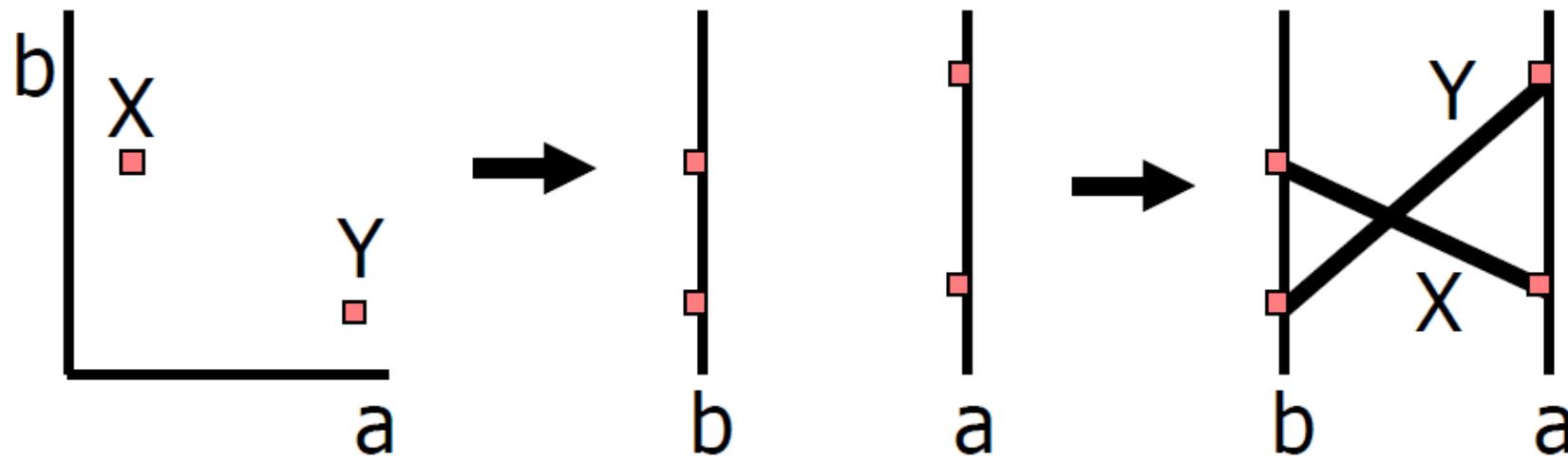


# Parallel Coordinate Plots



# Parallel Coordinate Plots

- All axes parallel
- Each sample is a line in this space
- Axes scaled to min/max range of data
- All dimensions can be shown at the same time
  - Positive / negative correlation == parallel / crossing lines

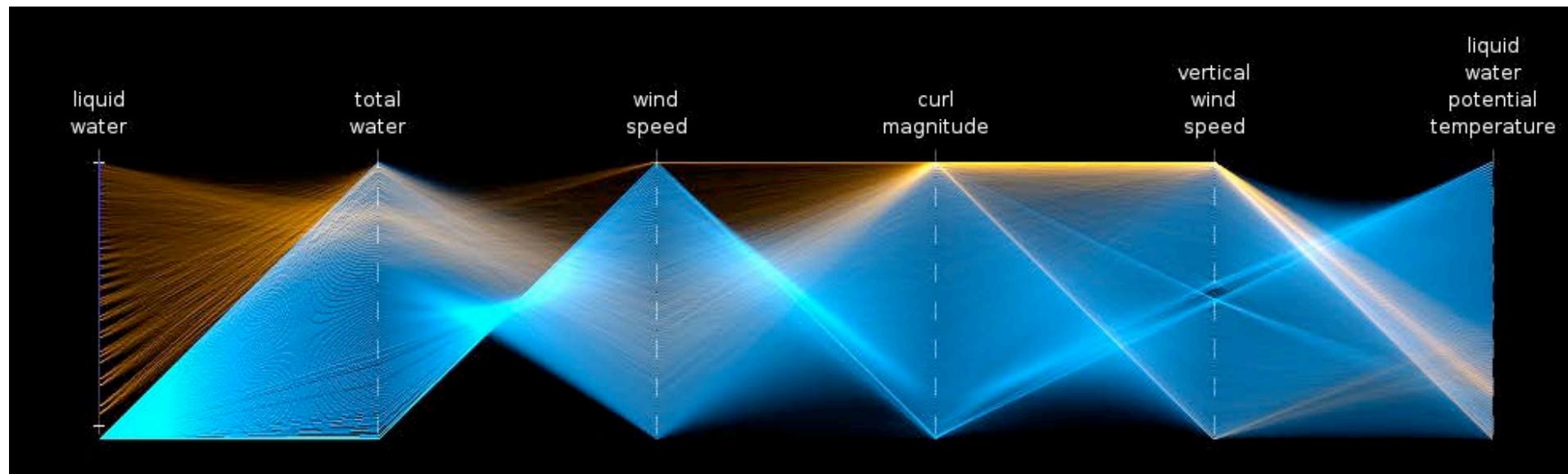


# Parallel Coordinate Plots

- Data:
  - Several keys, quantitative/categorical values
- Scalability:
  - Up to one dozen of keys
  - Hundreds of items (not thousands)
    - Larger sizes require special techniques, e.g. blending, or hierarchical approaches

# Parallel Coordinate Plots

- PCPs can and do scale up to millions of points, with some effort
  - E.g.: additive blending vs line drawing, histogram equalization



# Parallel Coordinate Plots

- Discussion:
  - Reordering dimensions can help finding patterns
    - Also interactive queries or filtering
  - They are relatively compact
    - Many variables can be shown simultaneously

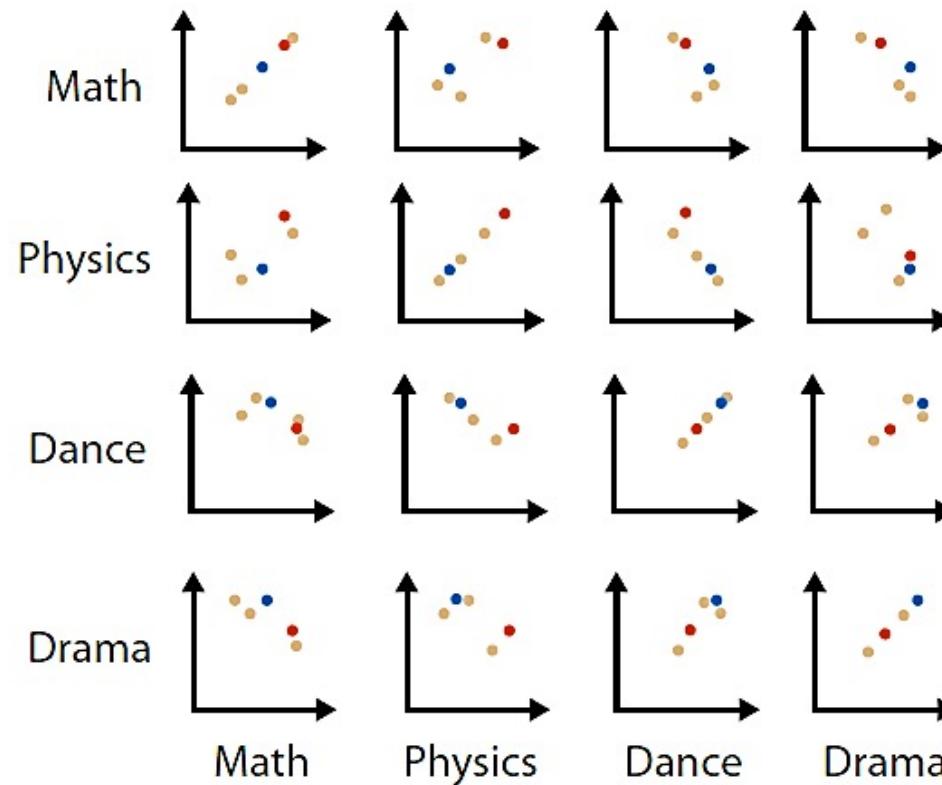
# Parallel Coordinate Plots

- Discussion. Relatively compact

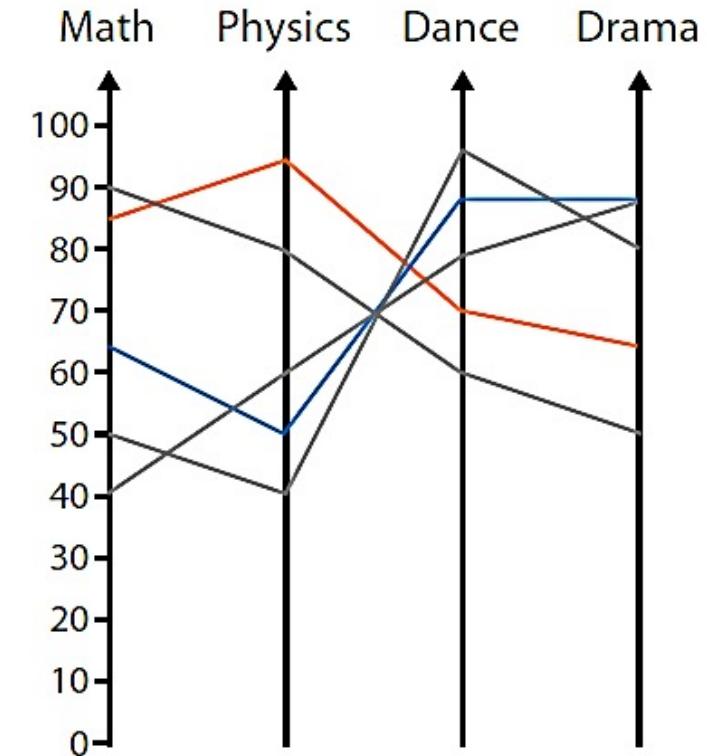
Table

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

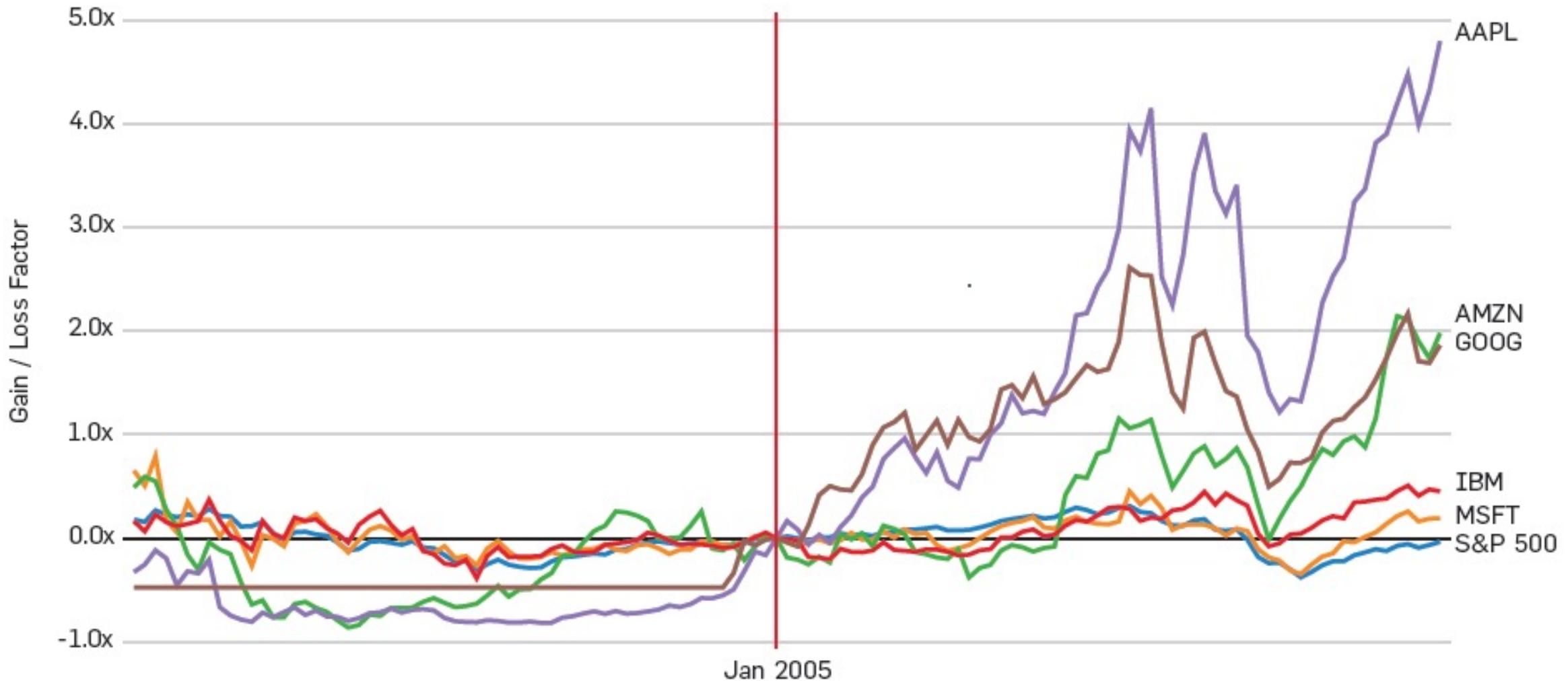
Scatterplot Matrix



Parallel Coordinates



# Time series. Index charts

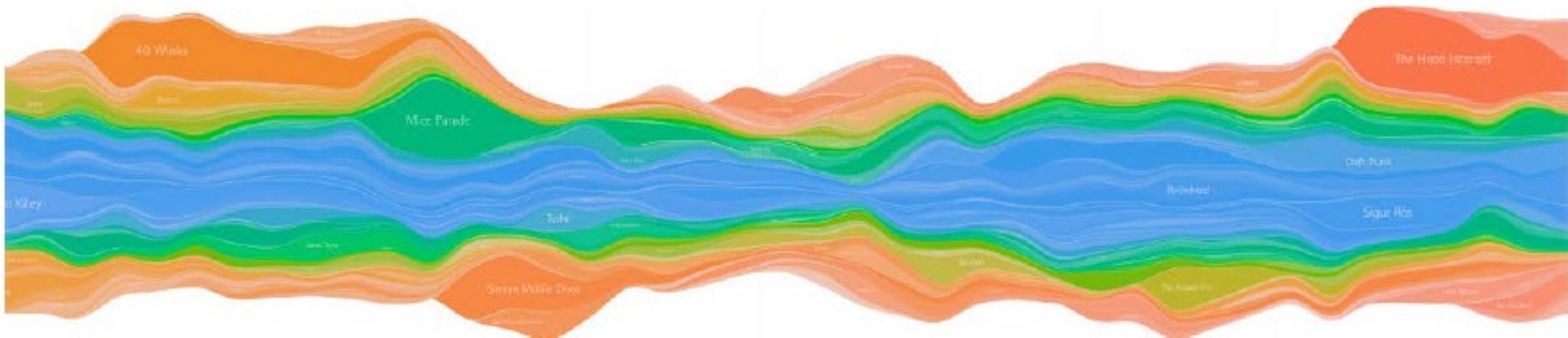


# Time series. Index charts

- Shows percentage changes for a collection of time-series data
  - Based on a selected index point
  - Suitable for displaying relative changes (e.g. stock market)
    - May lose context

# Time series. Streamgraph

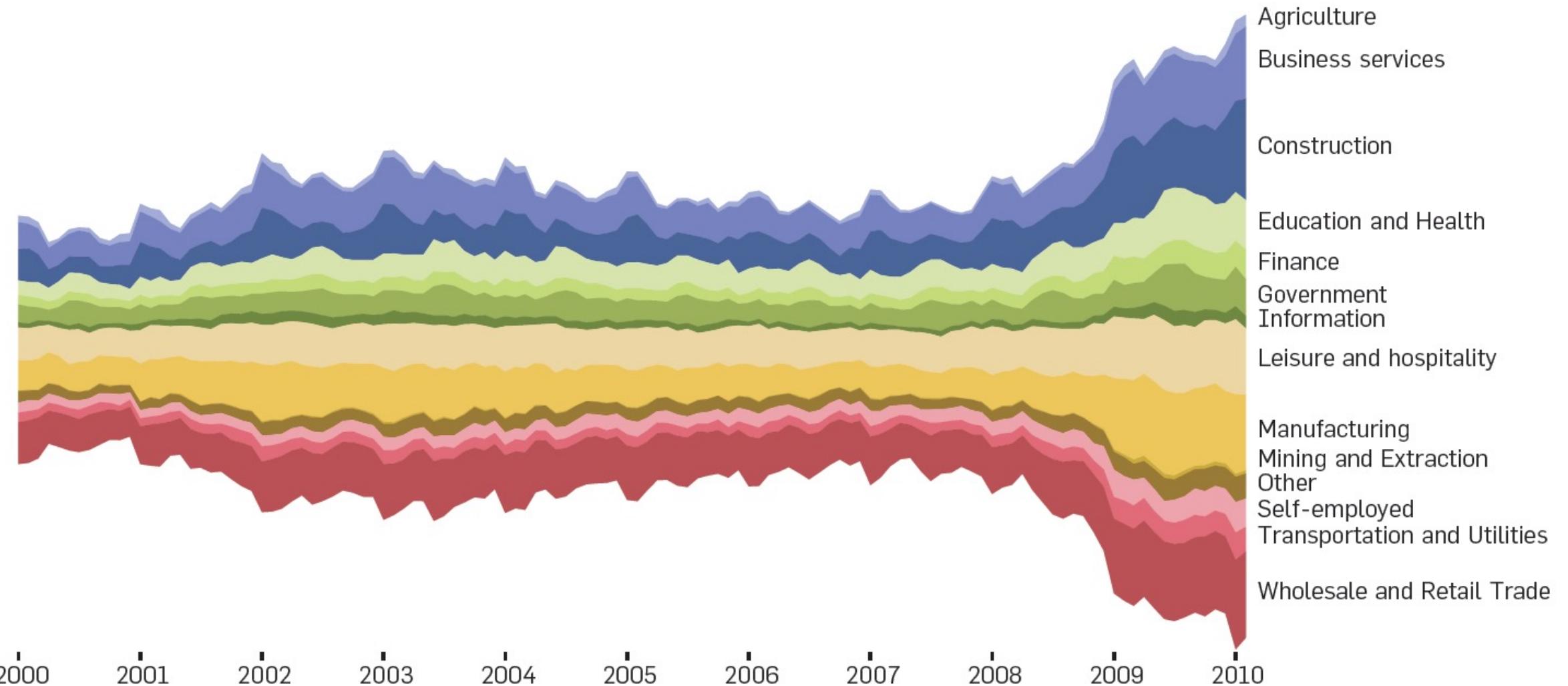
- Generalized stacked graph



# Time series. Streamgraph

- Streamgraph
  - Emphasizing horizontal continuity vs vertical items
  - Data
    - 1 categ key attrib (e.g. artist)
    - 1 ordered key attrib (e.g. time)
    - 1 quant value attrib (e.g. counts)
  - Derived data
    - geometry: layers, where height encodes counts
    - 1 quant attrib (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

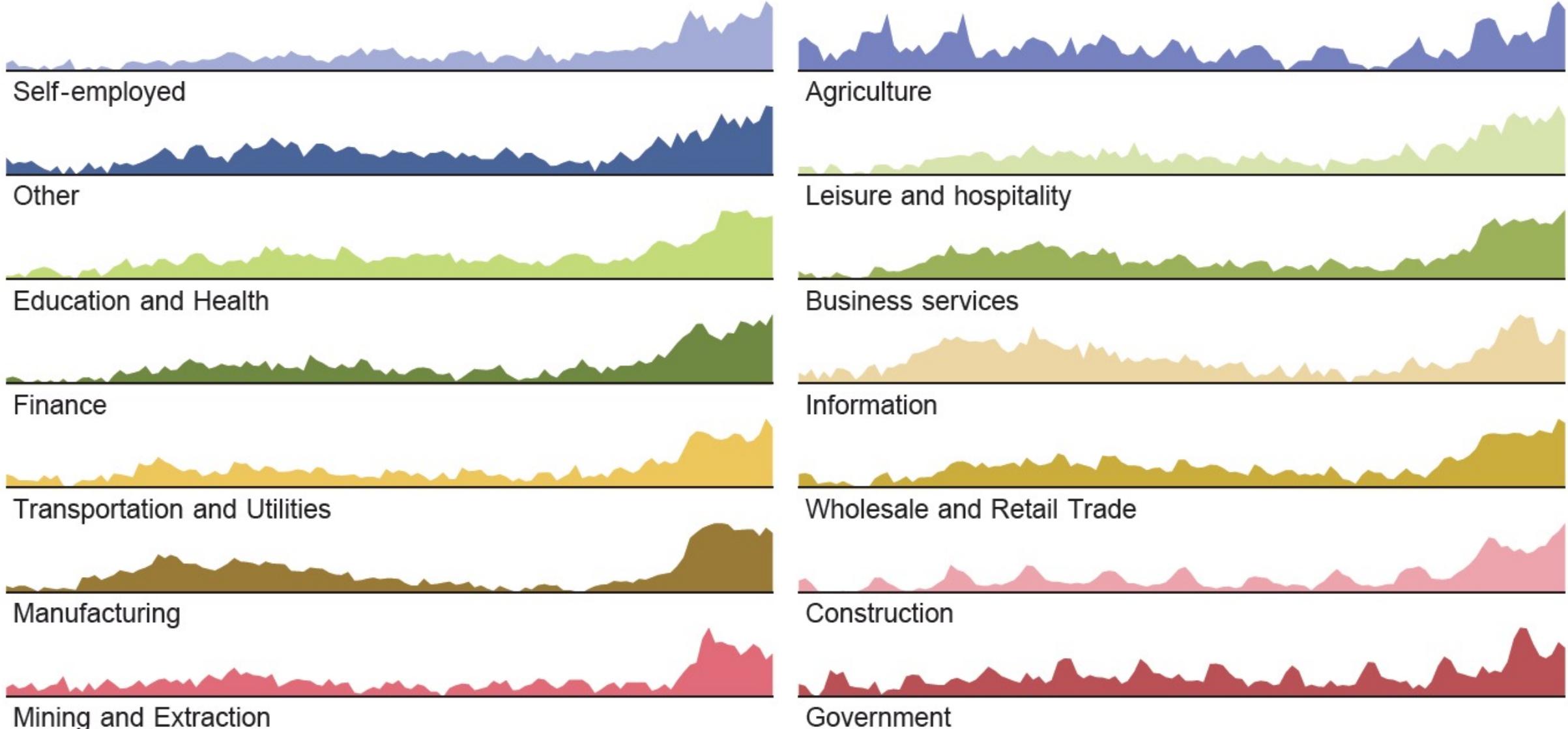
# Time series. Streamgraph



# Time series. Streamgraph

- Discussion:
  - Does not support negative values
  - Does not support data that cannot be added (e.g. temperatures)
  - Trends may be difficult to interpret
    - E.g. at the top of the curves

# Time series. Small multiples



# Time series. Small multiples

- Instead of stacking, placing series together
  - Same axes
- Task:
  - Find trends, seasonal patterns
- Can use any type of visualization

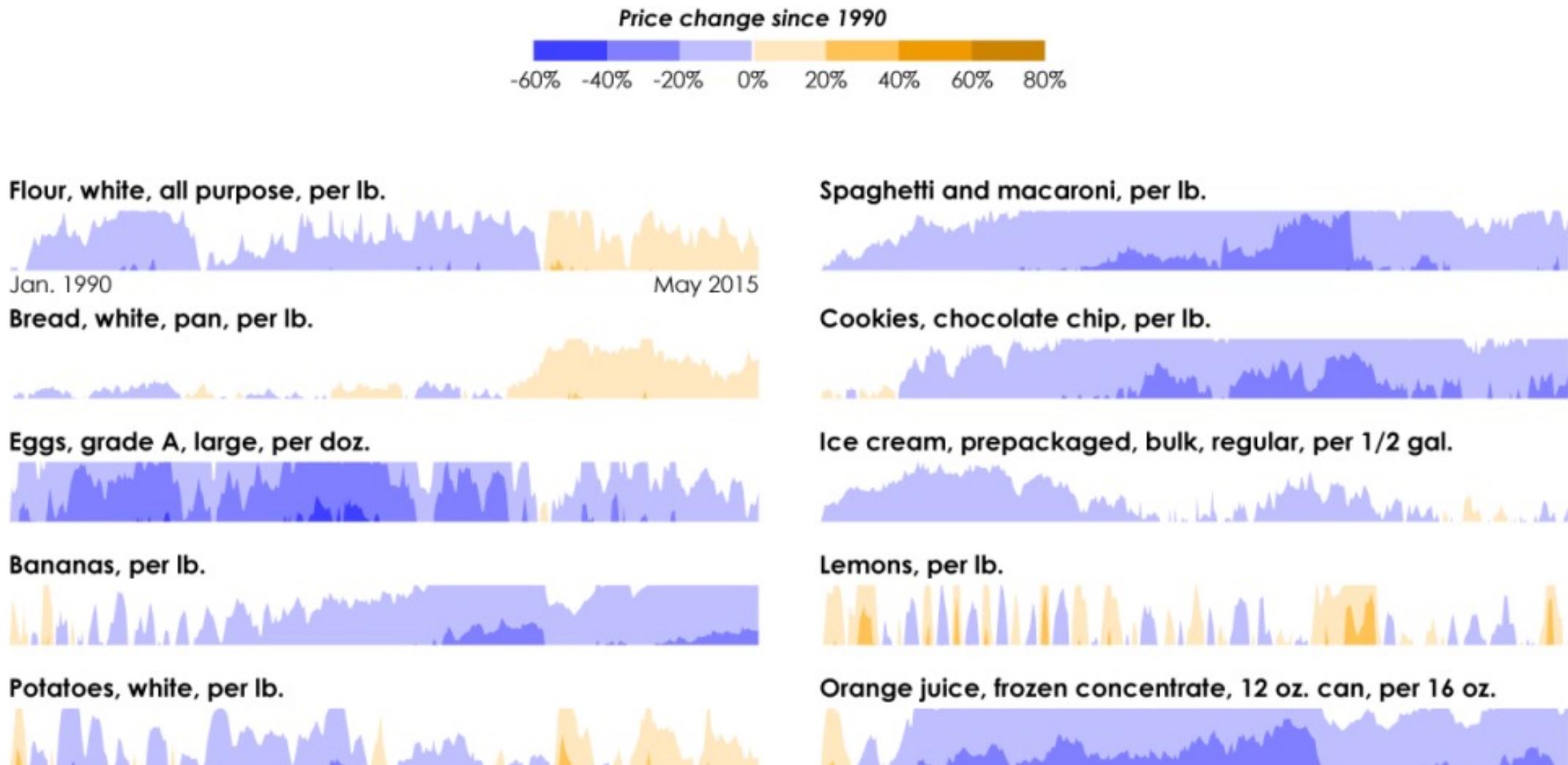
# Time series. Small multiples

2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support



Orange and green colors correspond to states where support for vouchers was greater or less than the national average.  
The seven ethnicreligious categories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.  
Where a category represents less than 1% of the voters of a state, the state is left blank.

# Representations. Horizon graphs

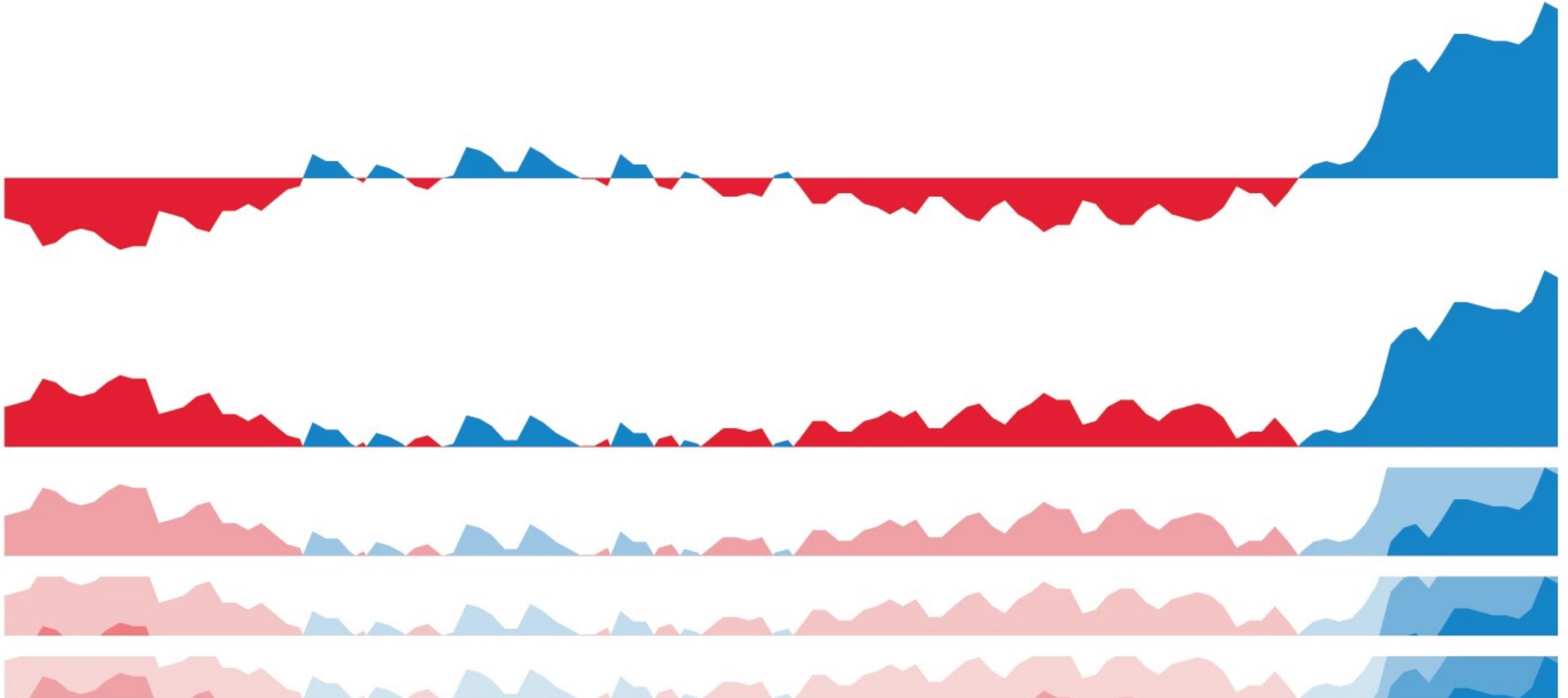


<https://flowingdata.com/2015/07/02/changing-price-of-food-items-and-horizon-graphs/>

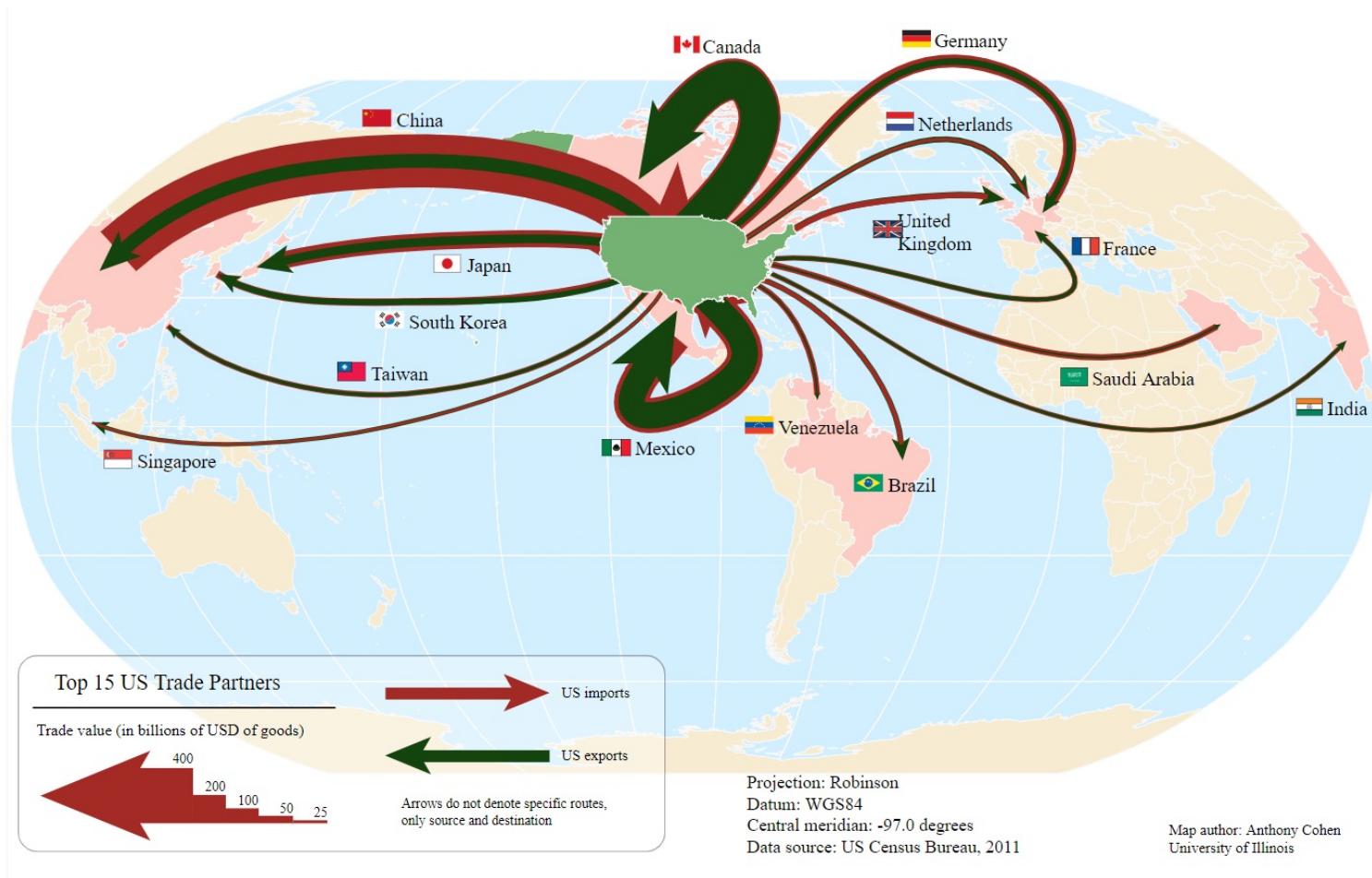
# Representations. Horizon graphs

- Increase the data density by overlapping
  - Keeping resolution
- Start with an area chart
  - Mirror negatives to the positive side
  - Divide the chart into bands, and mirror again
  - Divide the chart into bands, and mirror again
    - Result: 25% less of vertical space with same resolution
  - Takes time to learn

# Representations. Horizon graphs



# Representations. Flow maps



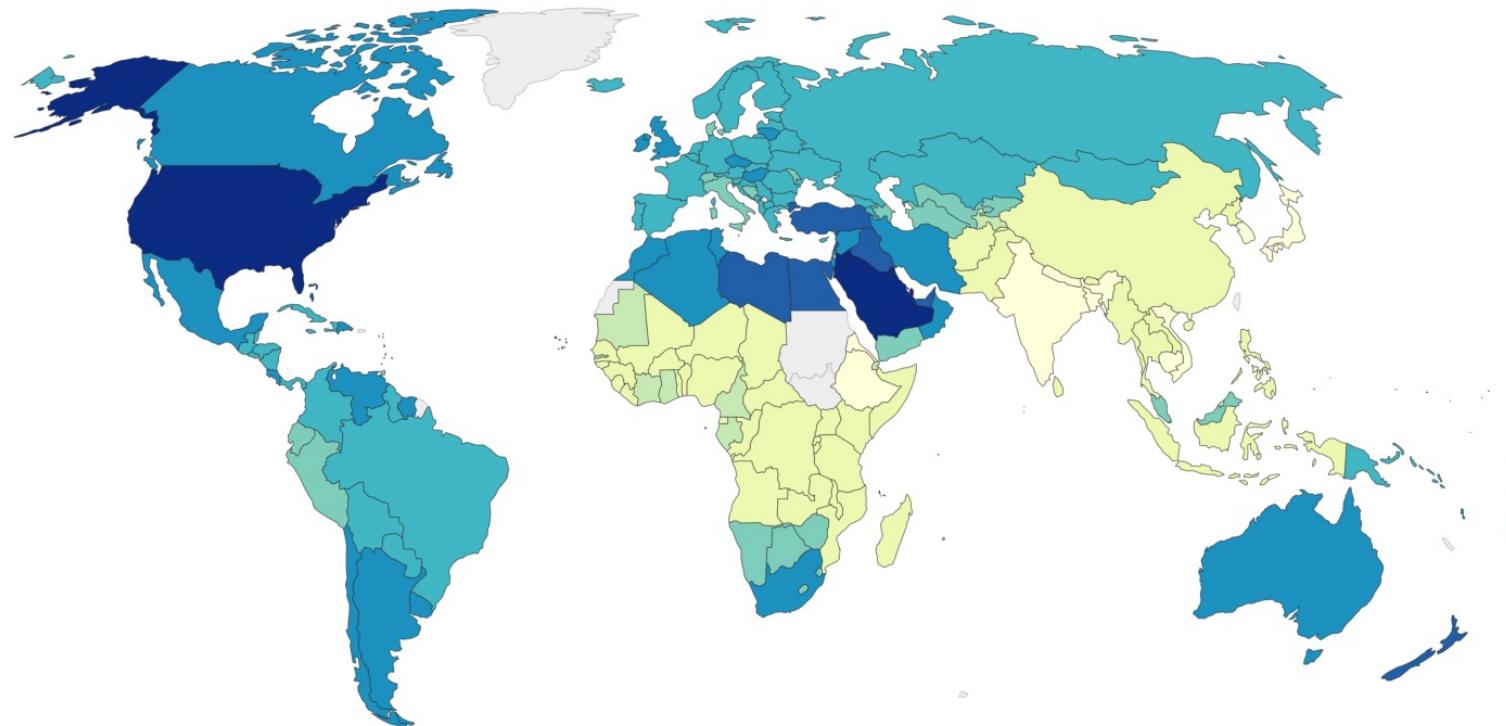
# Representations. Flow maps

- Depicts movement of a quantity in space
  - Implicitly in time
- Can encode a large amount of multivariate information
  - Path points, direction, line thickness, color...
- May require subtle distortion of the map

# Representations. Choropleth Maps

## Share of adults defined as obese, 2016

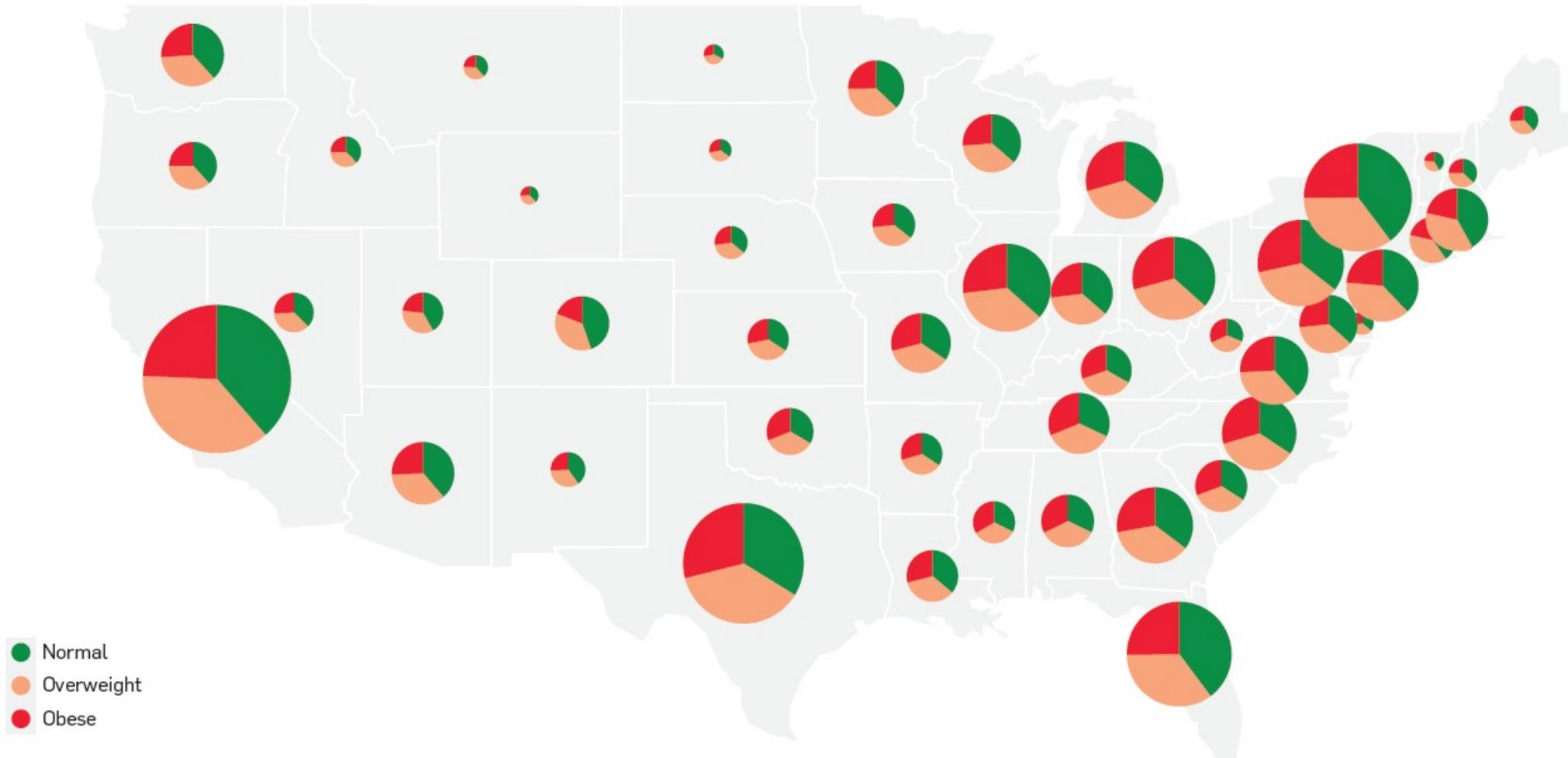
Percentage of adults aged 18+ years old who are defined as obese based on their body-mass index (BMI). BMI is a person's weight in kilograms (kg) divided by his or her height in metres squared. A BMI greater than or equal to 30 is defined as obese.



# Representations. Choropleth Maps

- Use colors to encode geographically aggregated data
- It is difficult to use properly:
  - Normalized data is often required (e.g. to communicate population density)
  - Shaded values can be perceived differently according to the area of the geographic region
- Encode: Color: Sequential segmented colormap

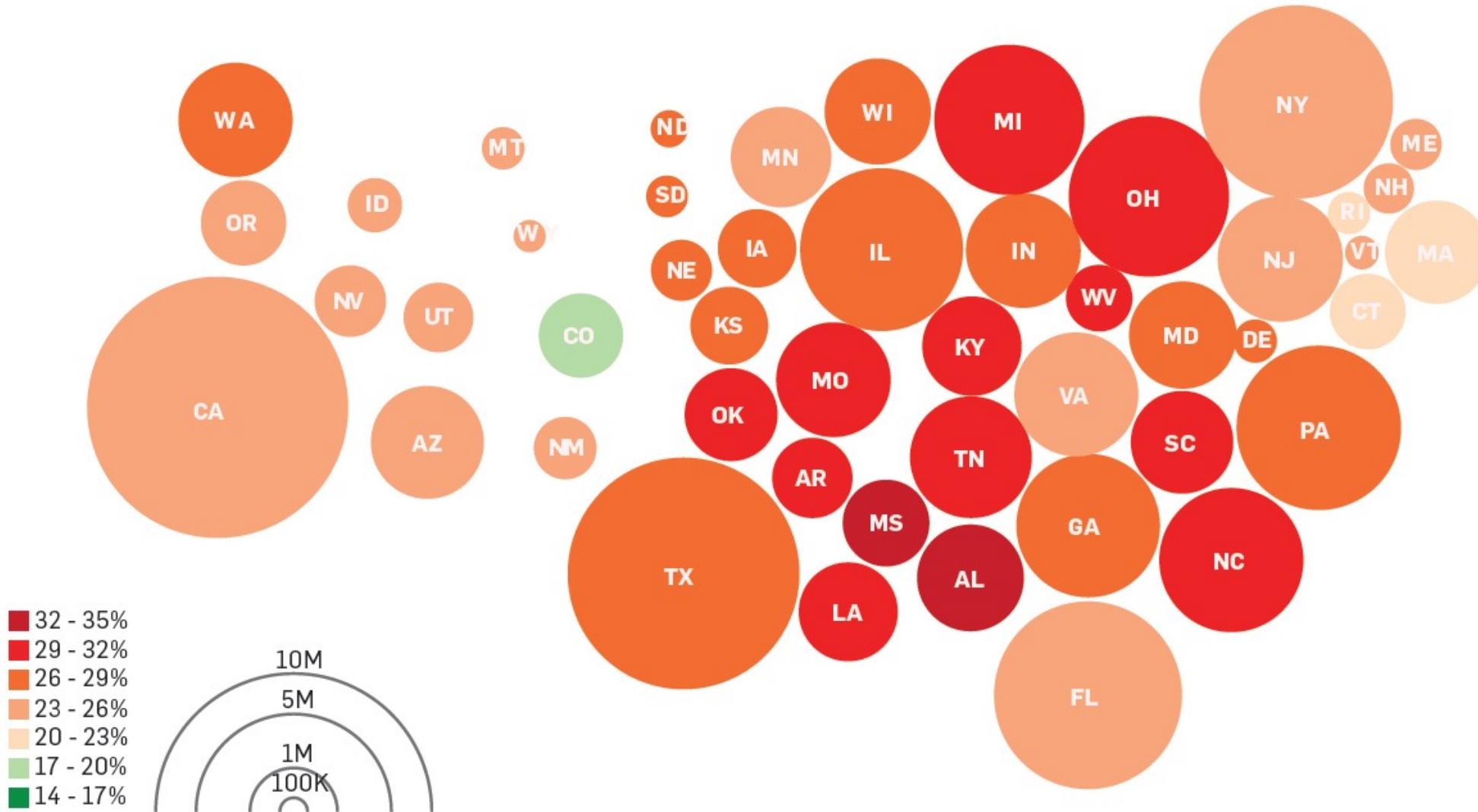
# Representations. Graduated symbol maps



# Representations. Graduated symbol maps

- Places symbols on the underlying map
- Avoids confounding geographic area with data values
- Enables visualizing more dimensions

# Representations. Cartograms

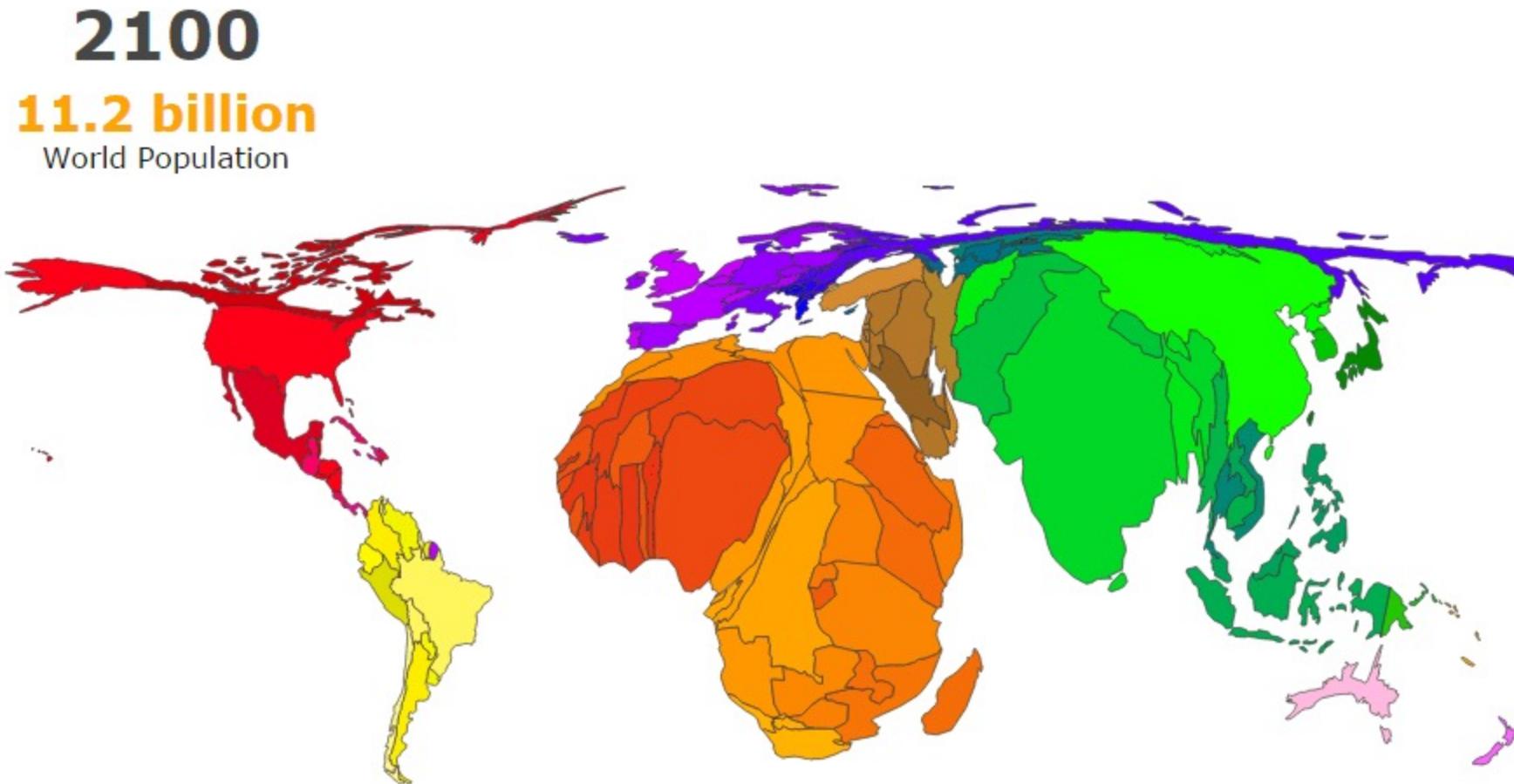


# Representations. Cartograms

- Distorts the shape of geographic regions to encode the data
- Dorling cartogram represents each geographic region with a circle

# Representations. Cartograms

- Many possibilities. E.g. population:



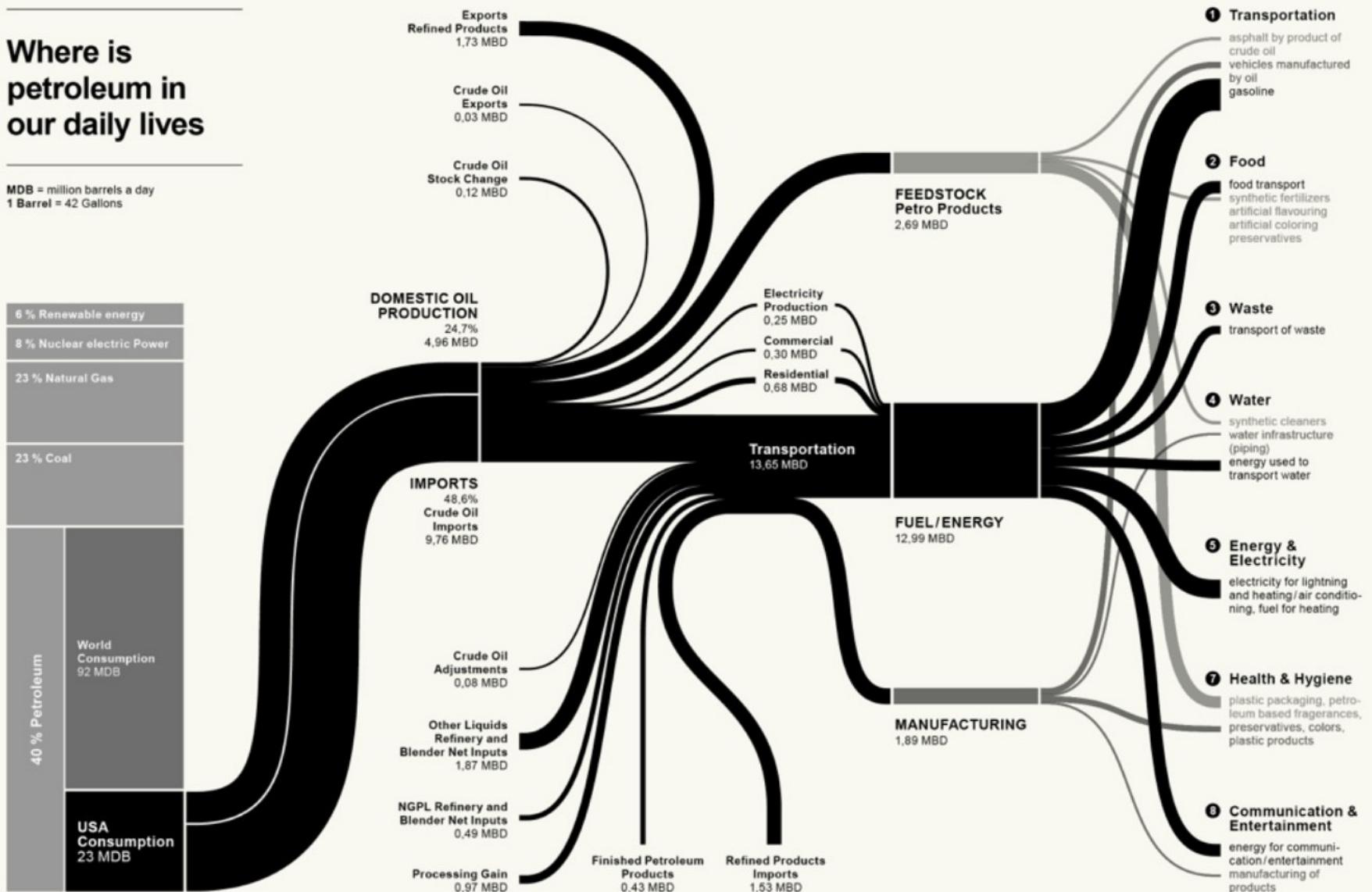
# Representations. Sankey diagrams

- Specific type of flow diagram
  - Width of the arrows are proportional to flow quantity
  - Put the emphasis on the major transfers of flows within a system
  - Useful to locate dominant contributions to an overall flow

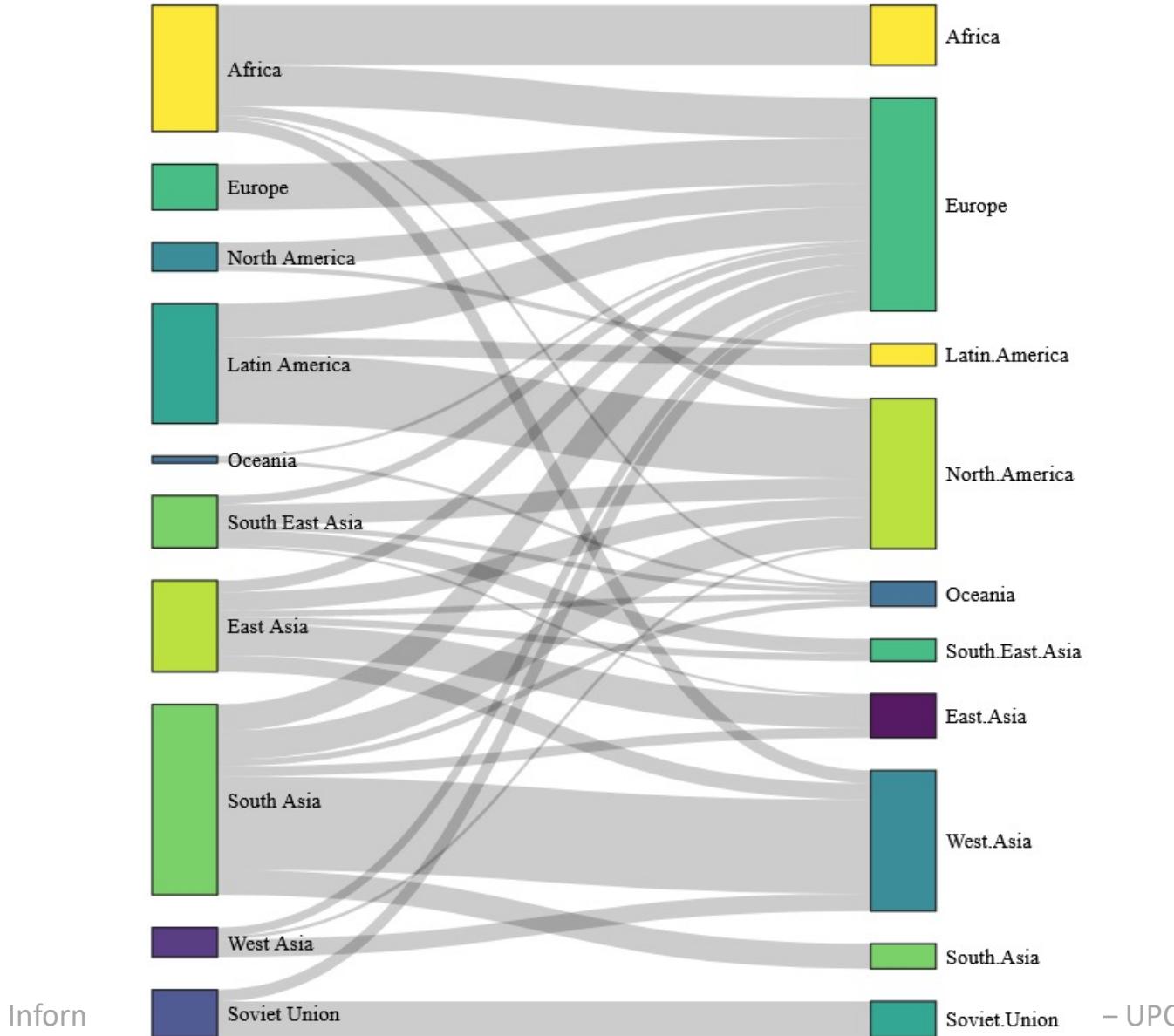
# Representations. Sankey diagrams

Where is  
petroleum in  
our daily lives

MDB = million barrels a day  
1 Barrel = 42 Gallons



# Representations. Sankey diagrams

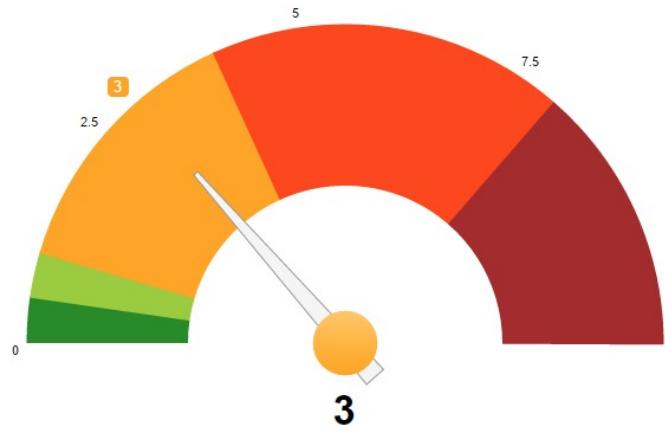


# Representations. Sankey diagrams

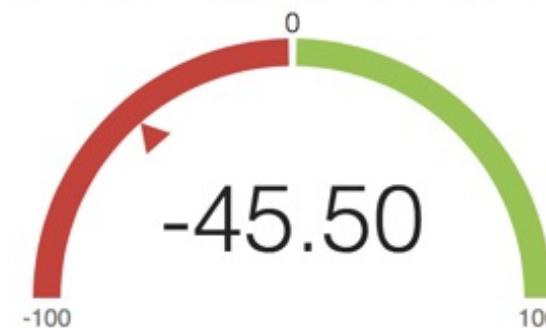
- Discussion:
  - Position of nodes is important
    - Minimize crossings
  - May make the figure unreadable if too many connections exist
    - Sometimes may dismiss weak connections

# Representations. Gauge & bullet charts

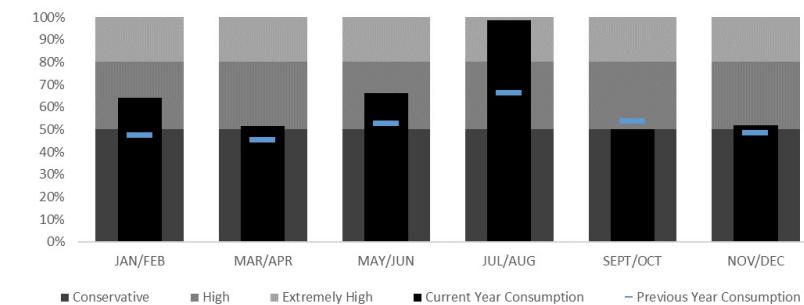
## Gauge charts



How satisfied are you with our product?



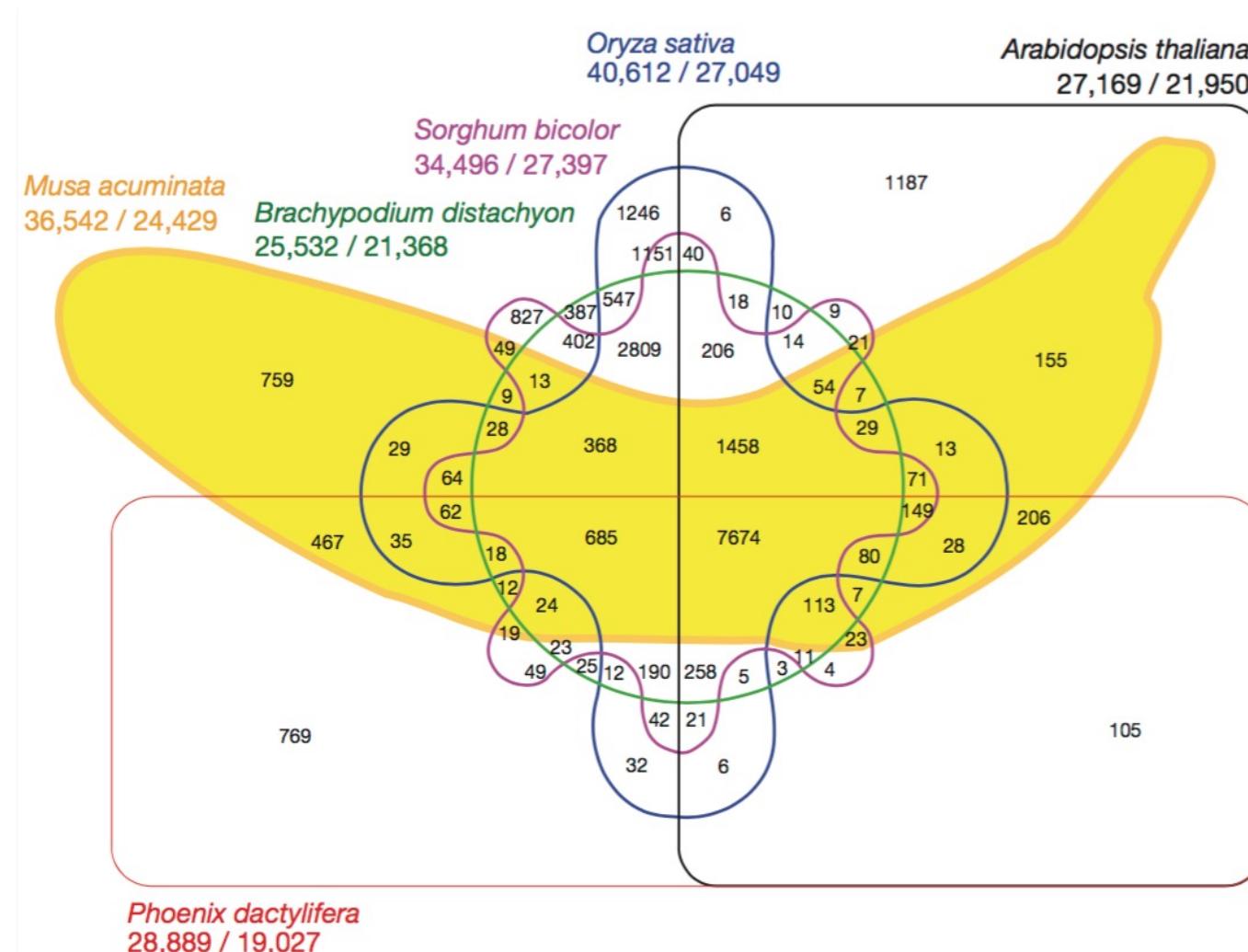
## Bullet charts



# Representations. Gauge & bullet charts

- Gauge charts:
  - Adaptation of real gauges
  - Current value (front) vs reference (background)
- Bullet charts:
  - Version of gauge charts using bars
  - Using the background of the bar chart to encode the reference value(s)

# Representations. Intersections of sets



Original Nature paper: <https://www.nature.com/articles/nature11241>

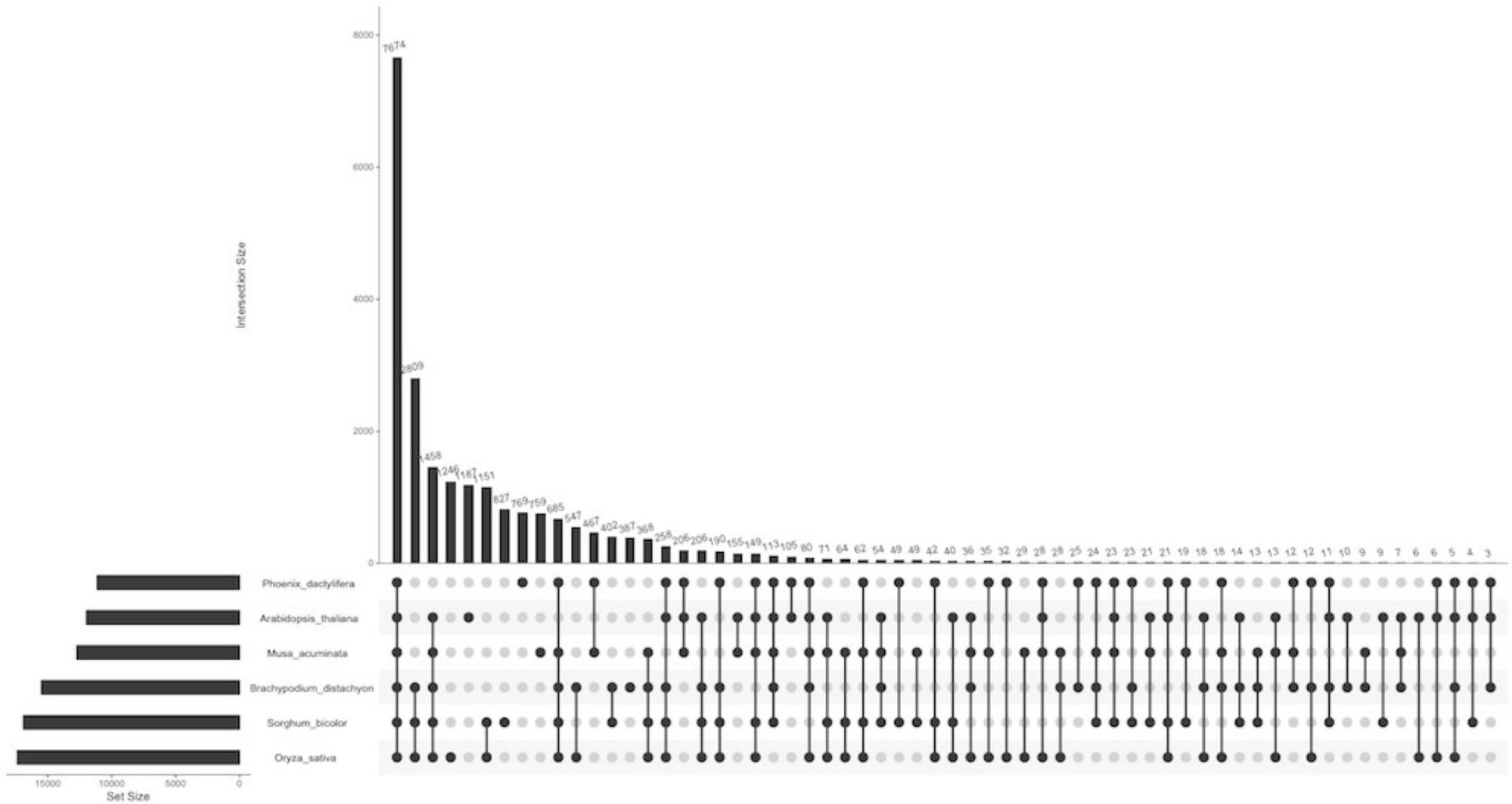
# Representations. Intersections of sets

- We will be tempted to visualize intersections through Venn diagrams
  - We know them
  - Seem intuitive
- Relationship between the banana's genome and the genome of five other species

# Representations. Intersections of sets

- The problem:
  - 3 sets have only 8 possible intersections
  - 10 sets have 1,024 possible intersections
  - $n$  sets yield  $2^n$  possible intersections !!!
- Using Venn (aka Euler) diagrams for more than 3-4 sets is a bad idea

# Representations. Upset diagram



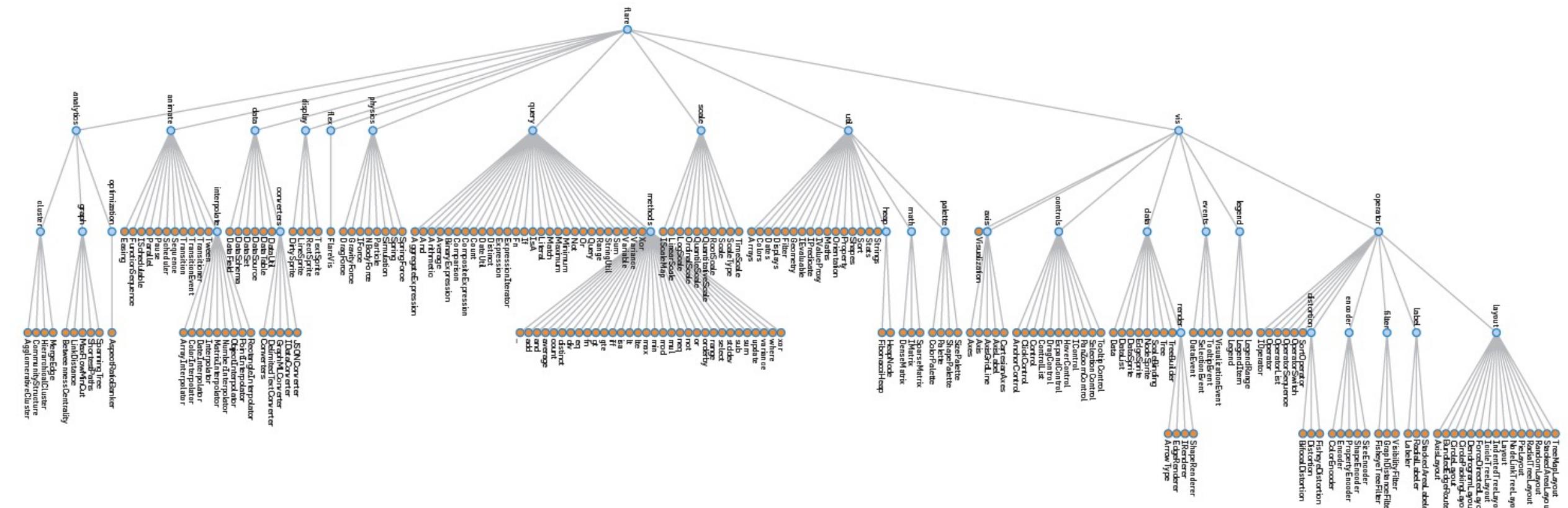
# Representations. Upset diagram

- Advantage:
  - Perceptually efficient visual encodings
  - Shows combinations of intersections
    - E.g., all the intersections involving two particular sets
  - Can add attributes about the intersections
    - Extra data associated with intersection can be encoded

# Representations. Hierarchical visualizations

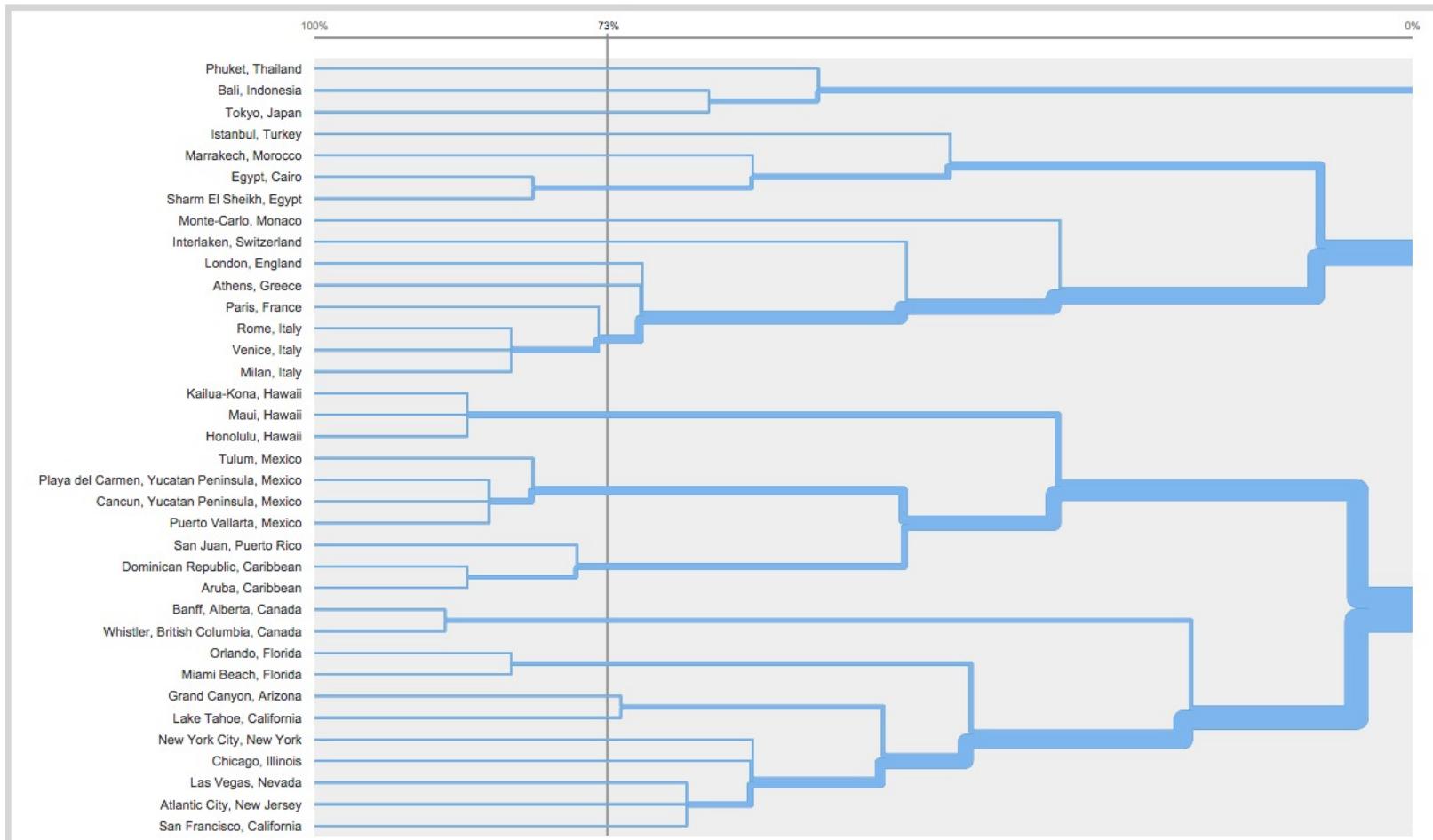
- Useful for visualizing hierarchies
- Tasks: finding hierarchical relationships, groups...

# Representations. Hierarchy. Node-link diagram

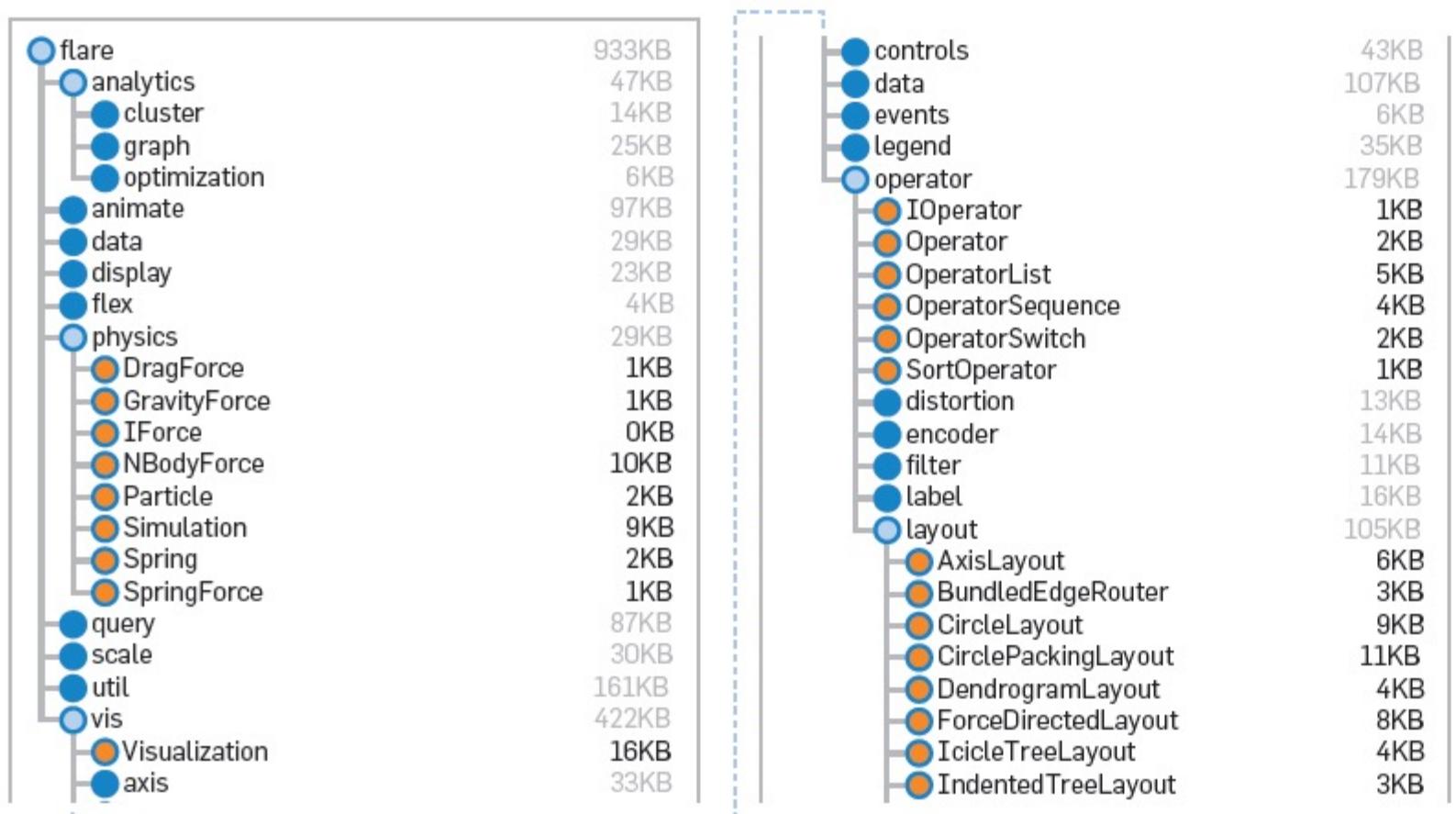


# Representations. Hierarchy. Dendrogram

- All the leaves at the same level (e.g. clustering algorithms)



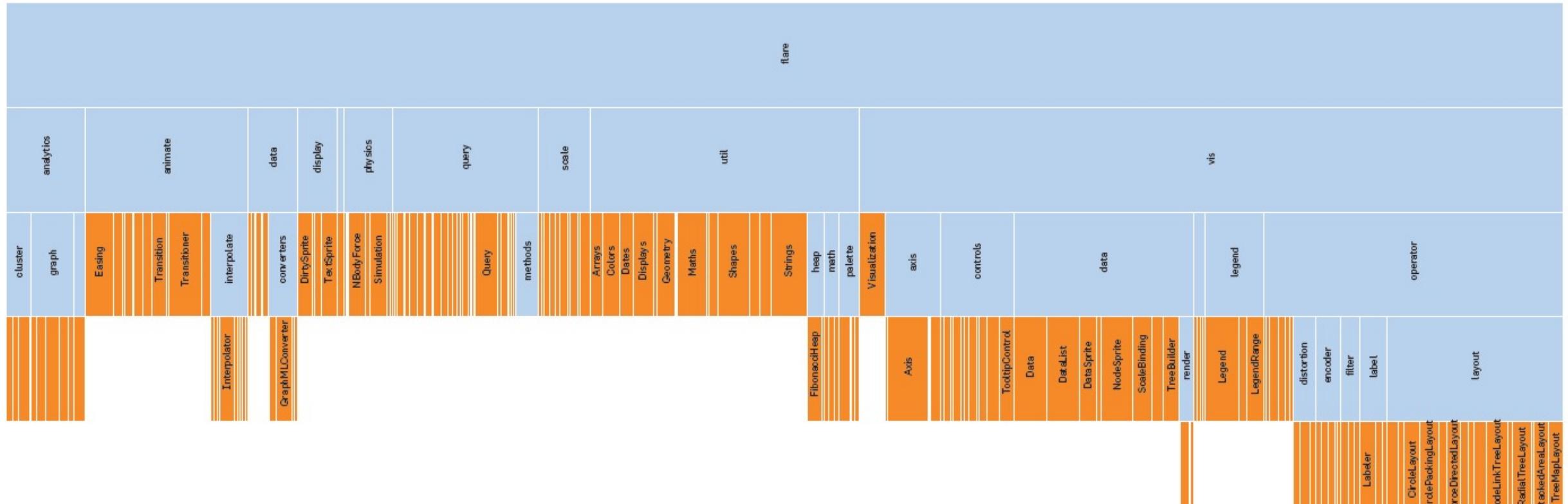
# Representations. Hierarchy. Indented tree



# Representations. Hierarchy. Indented tree

- Used in Operative Systems to depict the file directories
- Requires a large amount of vertical space
  - Does not facilitate multiscale inferences
- Allows for efficient interactive exploration of the tree
  - E.g. to find a specific node
- Allows for rapid scanning of node labels
- Multivariate data (e.g. file size, owner...) can be displayed adjacent to the hierarchy

# Representations. Hierarchy. Adjacency diagram

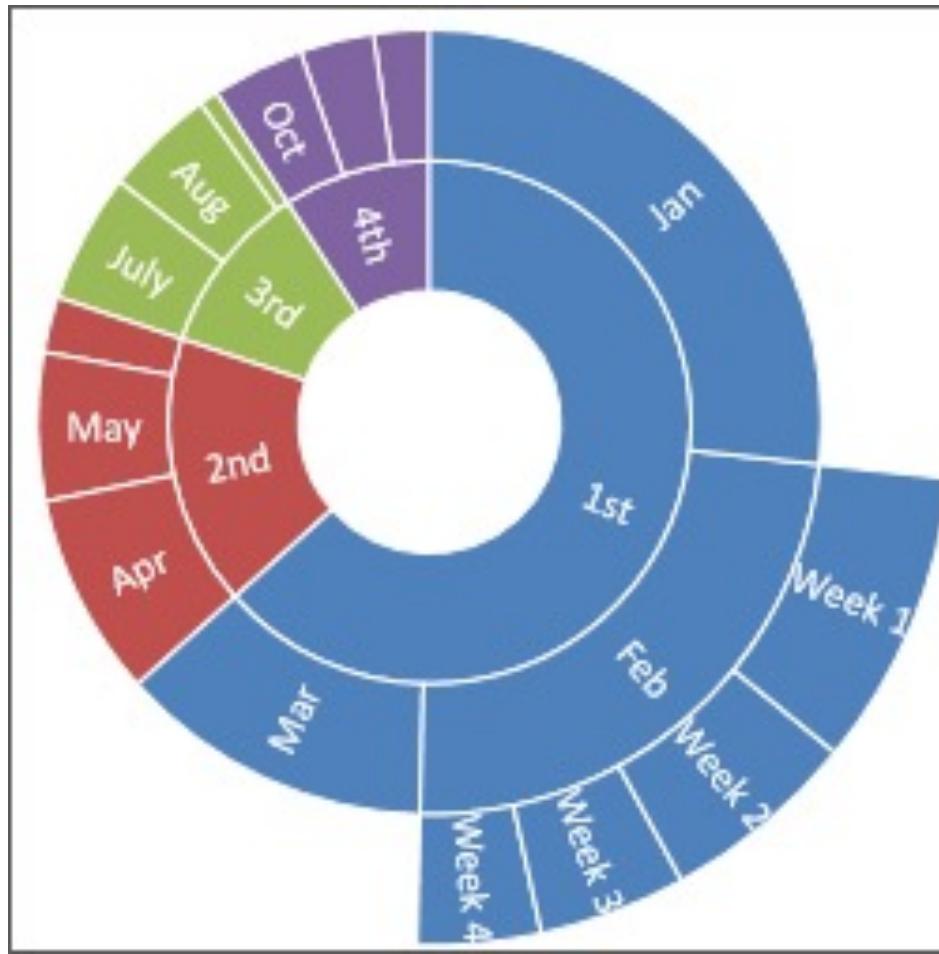


# Representations. Hierarchy. Adjacency diagram

- Space-filling variant of the node-link diagram
  - Nodes drawn as solid areas
  - Placement relative to adjacent nodes illustrates the position in the hierarchy
- Can use length to encode an additional dimension (e.g. size)

# Representations. Hierarchy. Sunburst

- Circular version of adjacency diagram



# Representations. Hierarchy. Treemap

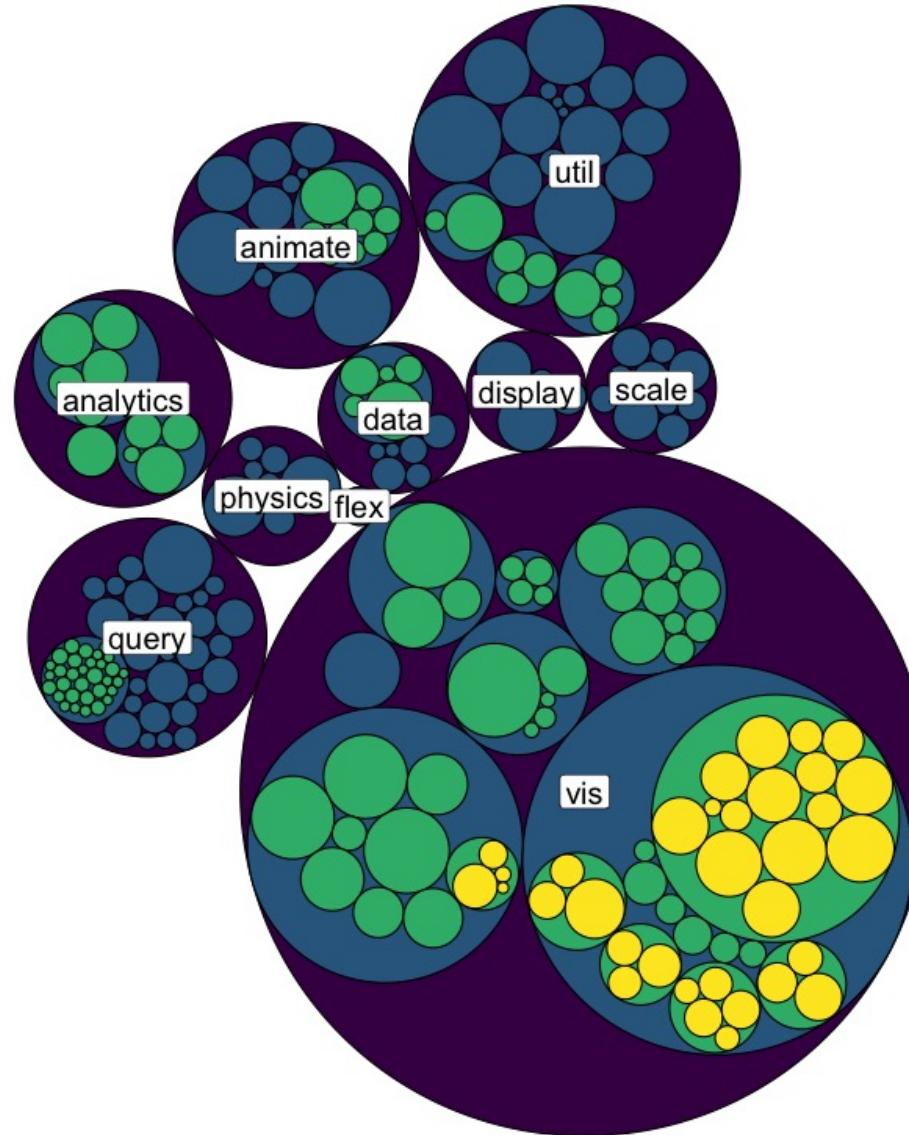
- AKA enclosure diagram



# Representations. Hierarchy. Treemap

- Space filling
  - Recursively subdivides the area into rectangles
- Squarified version uses approximately square rectangles
  - Typically better readability and size estimation
  - Can use circles

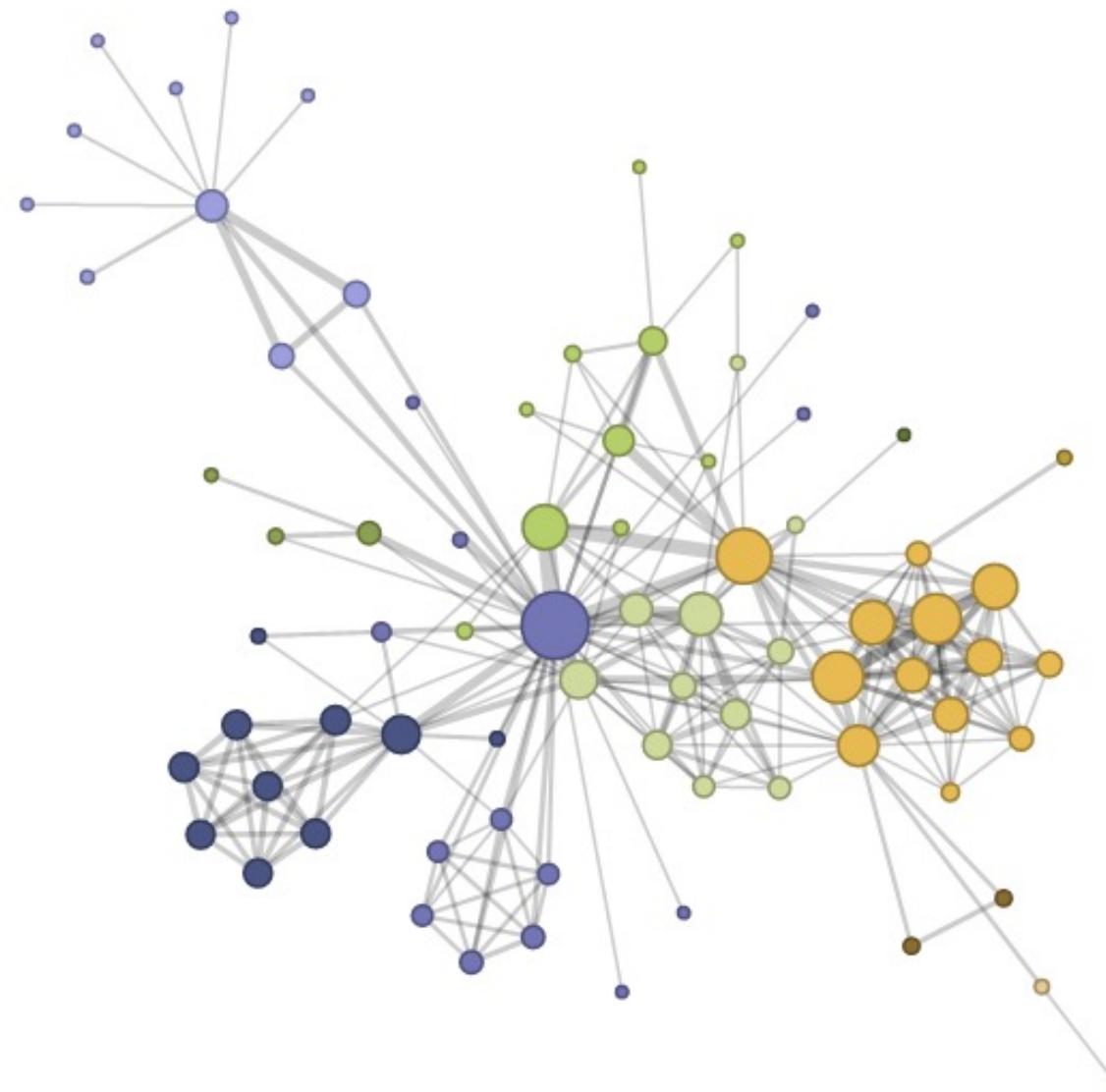
# Representations. Hierarchy. Treemap



# Representations. Networks

- Contain information about relationships
- Information such as: who is connected to, who is a central player (connected to many nodes), groups, cliques
- Drawing is highly complex
  - No given 2D position
  - Must place related nodes close and unrelated far away
  - Reduce crossings may facilitate legibility

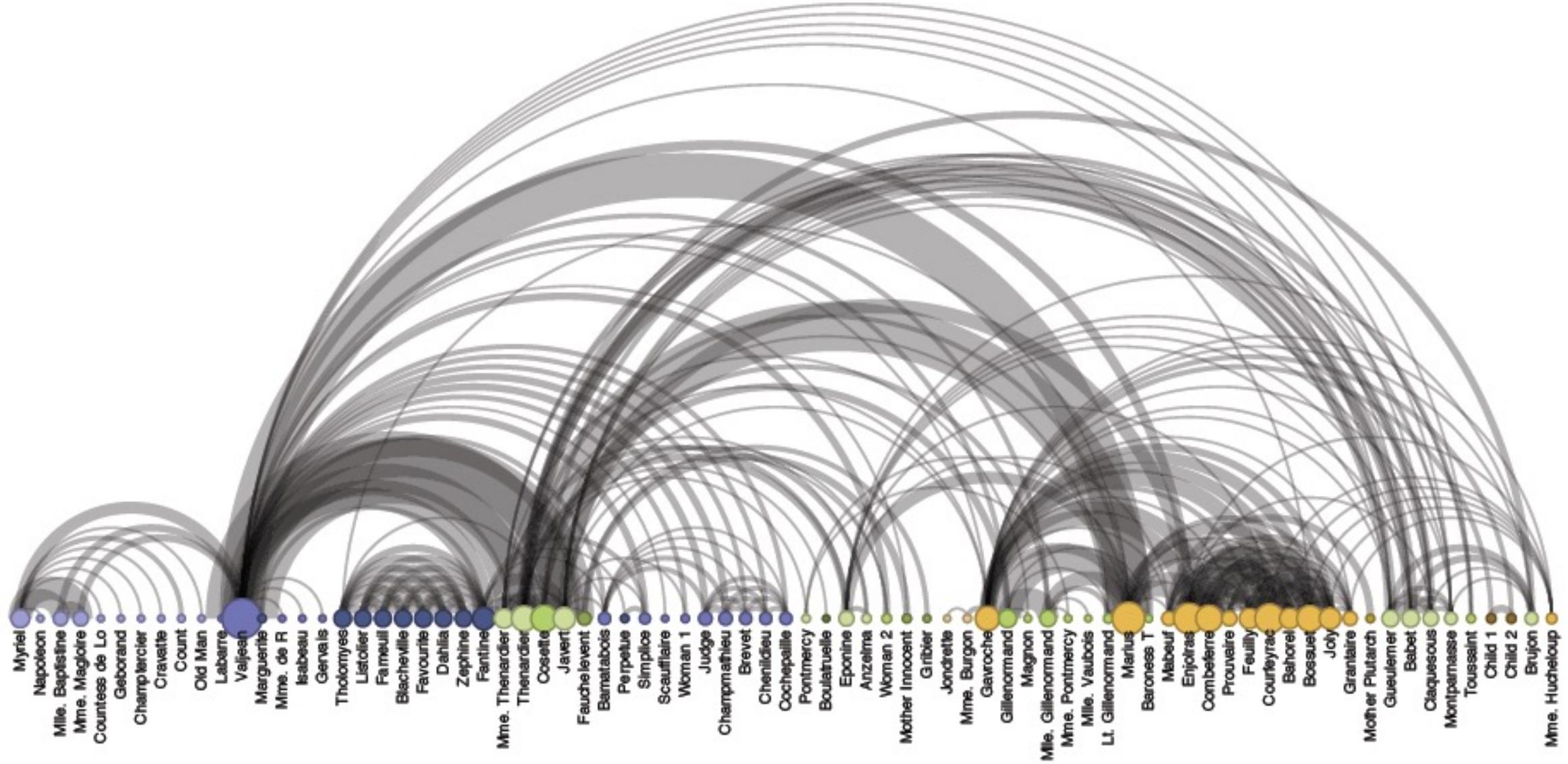
# Representations. Networks. Force-directed layout



# Representations. Networks. Force-directed layout

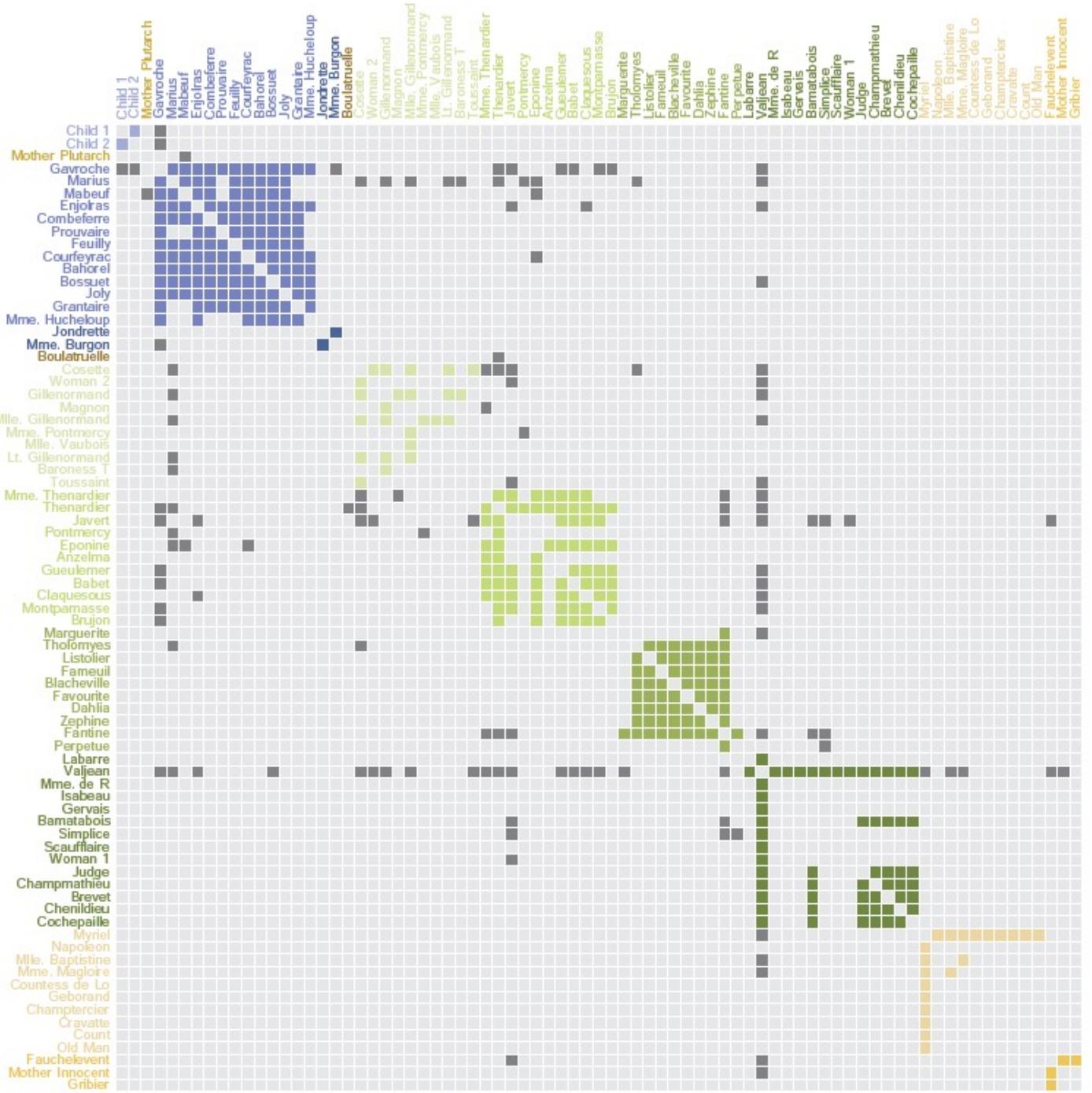
- Nodes are charged particles that repel each other
- Links are springs that pull related nodes together
- Uses physical simulation of the forces to determine node position
- Can add interaction to disambiguate links

# Representations. Networks. Arc diagram



# Representations. Networks. Arc diagram

- Lays the nodes in one dimension
- Circular arcs represent links
- Good ordering of nodes helps identifying cliques and bridges
- Multivariate data can be displayed alongside nodes
- Problems with the sorting of the data: seriation



# Representations. Networks. Adjacency matrix

- Nodes arranged in rows and columns
  - Intersections indicate links
- Using color or saturation instead of text facilitates identification of the links

# Representations. Networks. Adjacency matrix

- Discussion
  - Crossings are impossible
  - Effective sorting can reveal clusters and bridges
  - Interactive grouping and reordering facilitates deeper exploration of the network structure
  - Still problems with seriation of the data
  - More difficult to follow paths

# To know more...

- Jon Schwabish's YouTube series *One chart at a time*:  
[https://www.youtube.com/playlist?list=PLfv89tPxITiVlrwuSB\\_CISiBaGSH1CJR5-](https://www.youtube.com/playlist?list=PLfv89tPxITiVlrwuSB_CISiBaGSH1CJR5-)
- Do No Harm Guide:  
<https://www.urban.org/sites/default/files/publication/104296/do-no-harm-guide.pdf>

# Other visual tools

Pere-Pau Vázquez

Dept. Computer Science – UPC

# Multi-functioning elements

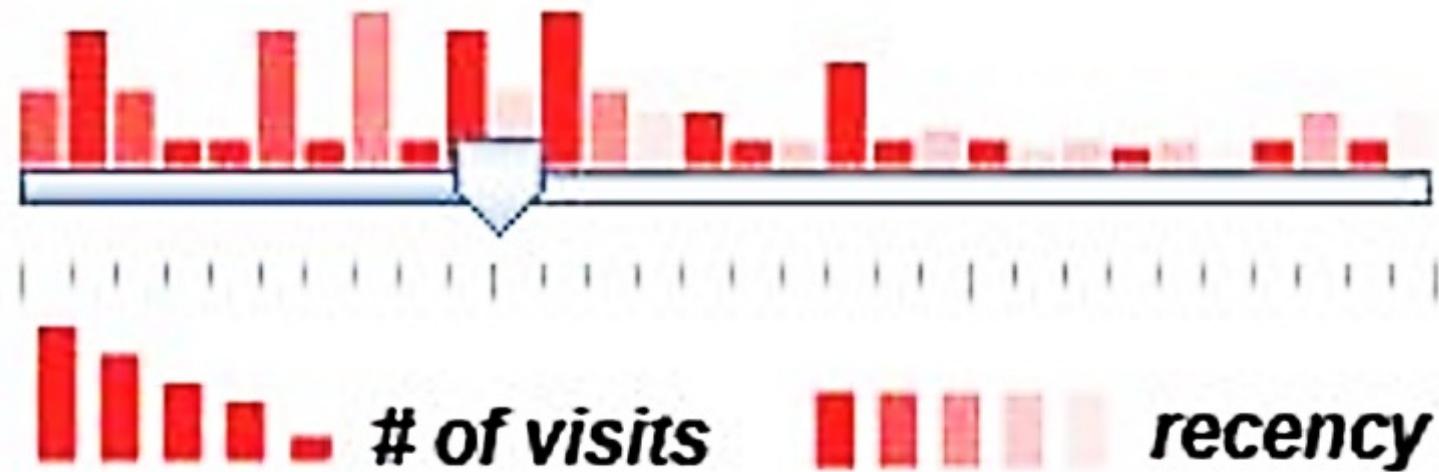
- Use the same ink for different purposes
- Maximizing data density:
  - Number of entries in data / area of data graphics

# Multi-functioning elements

- Recap: Signal to noise ratio (or ink-ratio):
  - Goal of communication: maximize signal and minimize noise
  - How?
    - Keep the design simple => avoid clutter, enhance perception
    - We can enhance information by using redundant coding and highlighting
    - Remove noise by eliminating unnecessary elements

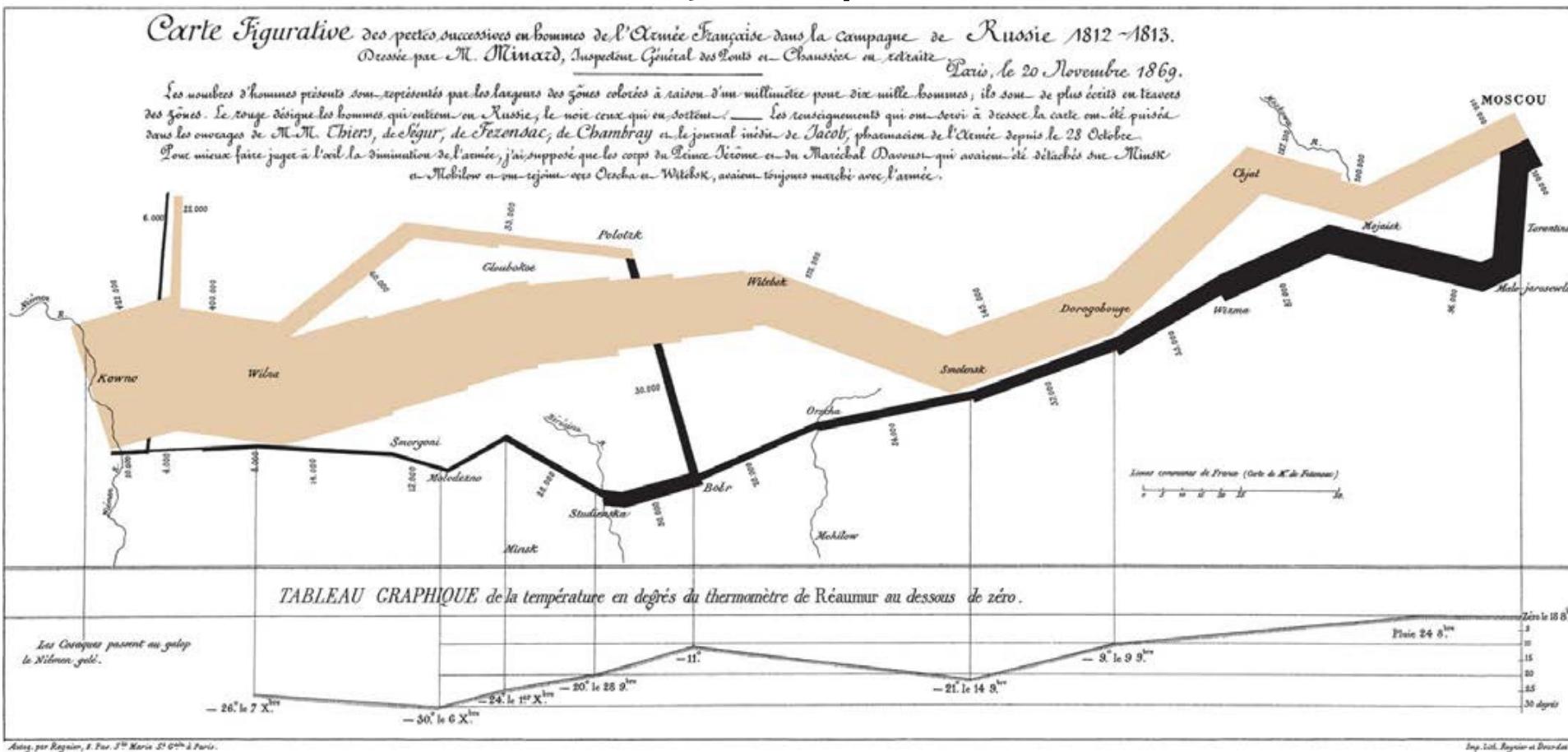
# Multi-functioning elements

- Scented widgets



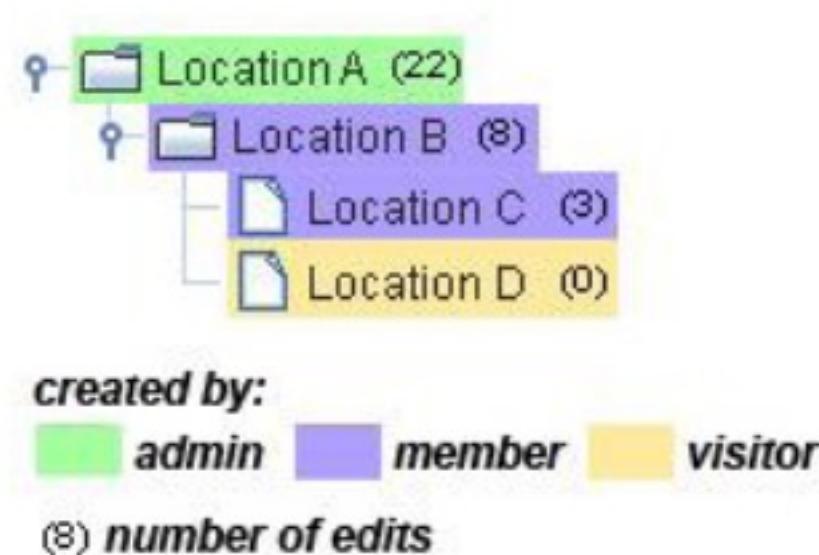
# Multi-functioning elements

- The dominant lines represent geographic location, size, and advance/retreat of the army troops



# Multi-functioning elements

- Same information in color and number, besides name



# Multi-functioning elements

- Sparklines

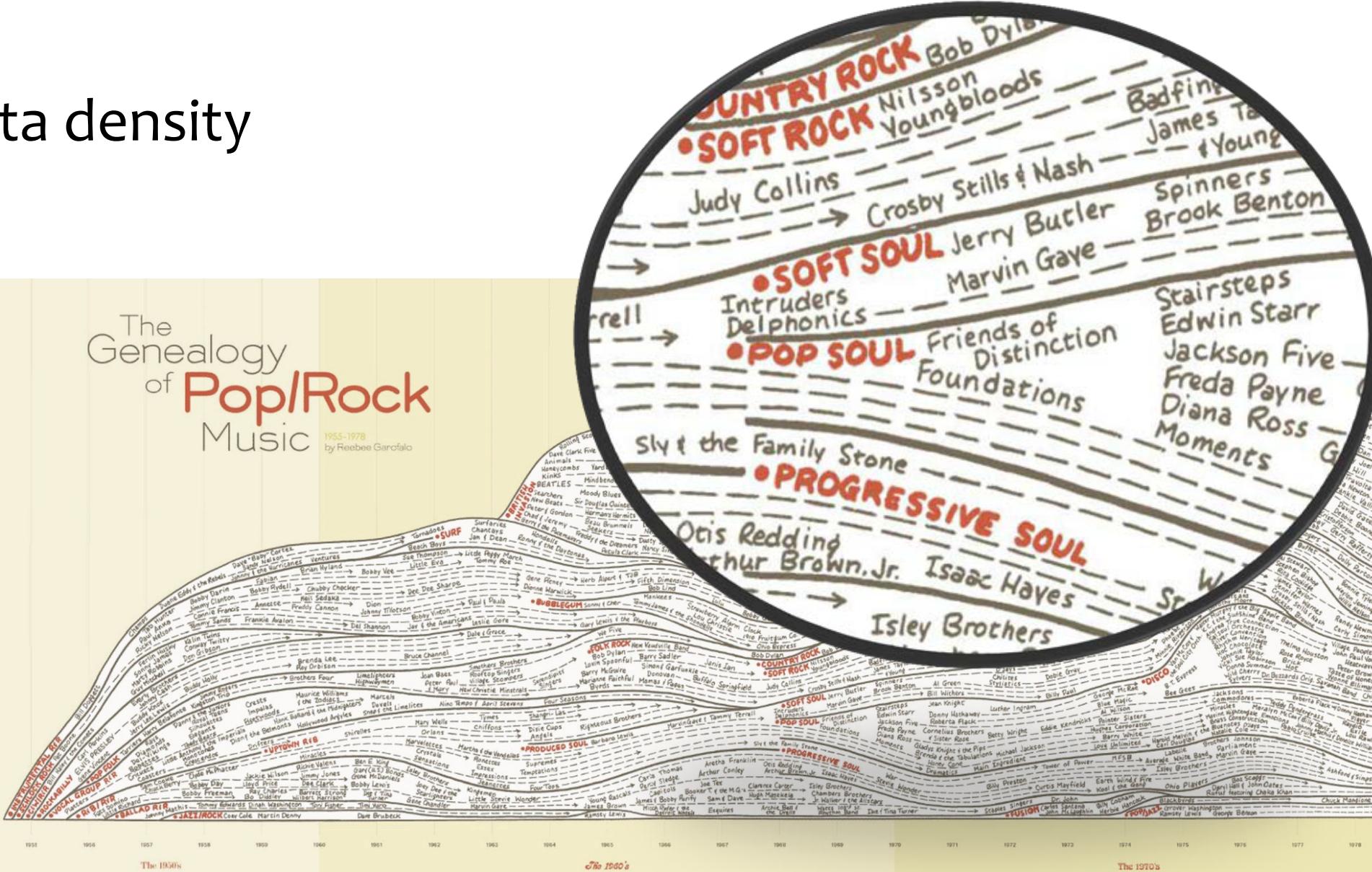
*Dequantification* In exchange for an enormous increase in graphical resolving power, the wordlike size of sparklines precludes the overt labels and scaling of conventional statistical displays. Most of our examples have, however, depicted *contextual methods* for quantifying sparklines: the gray bar for normal limits and the red encoding to link data points in sparklines to exact numbers  glucose 6.6 ; global scale bars and labels for sparkline clusters; and, probably best of all, surrounding a sparkline with an implicit data-scaling box formed by nearby numbers that label key data points (such as beginning/end, high/low)  1.1025  1.1907 1.0781 1.2858 . And now and then sparklines might be scaled by very small type:



*Production methods* Data lines produced by conventional statistical graphics programs must be gathered together, rescaled, and resized into sparklines. Sometimes this can be quickly done by cutting and pasting data lines, then resizing the printed output to sparkline resolutions. To produce and display really elegant sparklines, however, currently

# Multi-functioning elements

- Data density



## Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero, but it can be a target or a long-term average. Can also be used to show sentiment (positive/negative).

**Example FT uses**  
Trade surplus/deficit, climate change



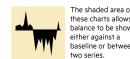
**Diverging stacked bar**



**Spine**



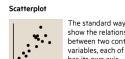
**Surplus/deficit filled line**



## Correlation

Show the relationship between two or more variables. It's important that, unless you have other evidence, readers will assume the relationships they see there to be causal (i.e. one causes the other).

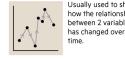
**Example FT uses**  
Inflation and unemployment, income and life expectancy



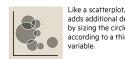
**Column + line timeline**



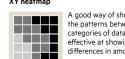
**Connected scatterplot**



**Ordered proportional symbol**



**Dot strip plot**



**Slope**



**Lollipop**



**Bump**



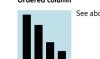
## Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

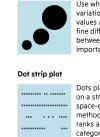
**Example FT uses**  
Wealth, deprivation, league tables, constituency election results



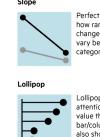
**Dot plot**



**Barcode plot**



**Slope**



**Violin plot**



**Population pyramid**



**Cumulative curve**



**Frequency polygons**



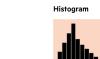
**Beeswarm**



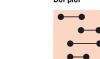
## Distribution

Show values in a dataset and how often they occur. The shape (skew) of distribution is a memorable way of highlighting the lack of uniformity or equality in the data.

**Example FT uses**  
Income distribution, population, Ogallala distribution, revealing inequality



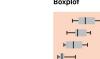
**Line**



**Column**



**Dot strip plot**



**Barcode plot**



**Slope**



**Area chart**



**Candlestick**



**Fan chart (projections)**



**Isotype (ictogram)**



**Connected scatterplot**



**Calendar heatmap**



**Priestley timeline**



**Circle timeline**



**Vertical timeline**



**Seismogram**



**Streamgraph**



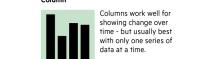
## Change over Time

Give emphasis to changing trends. These can be short (intraday) movements or longer periods traversing decades or centuries. Choosing the time period is important to provide suitable context for the reader.

**Example FT uses**  
Share price movements, economic time series, sectoral changes in a market



**Column**



**Bar**



**Paired column**



**Paired bar**



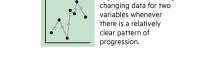
**Marimekko**



**Proportional symbol**



**Isotype (ictogram)**



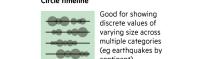
**Lollipop**



**Radar**



**Parallel coordinate**



**Circle timeline**



**Vertical timeline**



**Waterfall**



**Grouped symbol**



## Magnitude

Show size comparisons. These can be relative (just being able to see larger/smaller) or absolute (need to see fine differences). Usually these bars, bubbles, dollars or people rather than a calculated rate or per cent.

**Example FT uses**  
Fiscal budgets, company structures, national election results



**Marimekko**



**Pie**



**Donut**



**Treemap**



**Voronoi**



**Scaled cartogram (value)**



**Dot density**



**Gridplot**



**Venn**



**Waterfall**



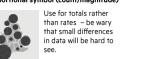
## Spatial

Show how a single entity can be broken down into its component elements. If the component is small in size of the entity in data are needed to represent the components.

**Example FT uses**  
Production density, natural resource locations, disaster risk/impact, catchment areas, variation in election results



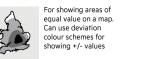
**Proportional symbol (count/magnitude)**



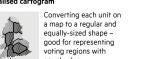
**Waterfall**



**Chord**



**Network**



## Flow

Show the reader volumes or intensity of movement between two or more states or locations. These might indicate sources or destinations of geographical locations.

**Example FT uses**  
Movement of funds, trade, migrants, laws, information, relationship graphs.



**Sankey**



# Visual vocabulary

## Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, no a wizard, but is a useful starting point for making informative and meaningful data visualisations.



[ft.com/vocabulary](http://ft.com/vocabulary)

# FT

© Financial Times

# Further reading

- <https://datavizproject.com/data-type/>
- <https://public.tableau.com/es-es/s/gallery/visual-vocabulary>
- <https://github.com/ft-interactive/chart-doctor/tree/master/visual-vocabulary>
- <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>
- Timeline of datavis milestones  
<http://www.datavis.ca/milestones/>

# Acknowledgments

- For these slides, I've used slides/papers by A. Vilanova, T. Munzner, J. Heer, and others... Thanks to all of them!

# Advanced Visualization Techniques

Pere-Pau Vázquez

Dept. Computer Science – UPC