

Visualization. Multi- Dimensional Data Visualization

Pere-Pau Vázquez
Dept. Computer Science – UPC

Outline

- Data sets
- Multi-Dimensional Data
- Evaluating layouts

Outline

- **Data sets**
- Multi-Dimensional Data
- Evaluating layouts

Data sets

- Data comes in many different forms
 - Typically, not in the way you want it
 - Heterogeneous data often seen as multiple dimensions of elements extracted by patterns or needs

Data sets

- Basics

164	dodge-st-rep	16.0	8.	400.0	170.0	4668.	11.5	75.	1.	
165	fiat-124-spor	15.0	8.	350.0	145.0	4440.	14.0	75.	1.	
166	fiat-124-tc	16.0	8.	318.0	150.0	4498.	14.5	75.	1.	
167	fiat-124b	14.0	8.	351.0	148.0	4657.	13.5	75.	1.	
168	fiat-128	17.0	6.	231.0	110.0	3907.	21.0	75.	1.	
169	fiat-128	16.0	6.	250.0	105.0	3897.	18.5	75.	1.	
170	fiat-131	15.0	6.	258.0	110.0	3730.	19.0	75.	1.	
171	fiat-strada-cl	18.0	6.	225.0	95.00	3785.	19.0	75.	1.	
172	fiat-x1.9	21.0	6.	231.0	110.0	3039.	15.0	75.	1.	
173	ford-country	20.0	8.	262.0	110.0	3221.	13.5	75.	1.	
174	ford-country	13.0	8.	302.0	129.0	3169.	12.0	75.	1.	
175	ford-country	29.0	4.	97.00	75.00	2171.	16.0	75.	3.	
176	ford-escort-2	23.0	4.	140.0	83.00	2639.	17.0	75.	1.	
177	ford-escort-4	20.0	6.	232.0	100.0	2914.	16.0	75.	1.	
178	ford-f108	23.0	4.	140.0	78.00	2592.	18.5	75.	1.	
179	ford-f250	24.0	4.	134.0	96.00	2702.	13.5	75.	3.	
180	ford-fairmon	25.0	4.	90.00	71.00	2223.	16.5	75.	2.	
181	ford-fairmon	24.0	4.	119.0	97.00	2545.	17.0	75.	3.	
182	ford-fairmon	18.0	6.	171.0	97.00	2984.	14.5	75.	1.	

Data sets

- Complex, multi-dimensional data is everywhere
 - E.g. Public datasets

After selecting variables, please click one of the above options

DIMENSION FILTERS

Hierarchy

[Collapse all](#) | [Expand all](#) | [Unselect all](#)

- ▶ Income
- ▶ Lending
- ▶ Region

▶ Aggregates

▶ COUNTRY

(Available: 214) Selected: 1

▶ SERIES

(Available: 1300) Selected: 0

▶ TIME

(Available:) Selected: 0

YOUR CURRENT SELECTION

▼ DATABASE

World Development Indicators

[Change database](#)

▼ COUNTRY (1)

[Remove all](#) | [Sort](#)

Drag to rearrange the order

 France

▶ SERIES (0)

▶ TIME (0)

214 countries



Afghanistan	Albania	Algeria	Andorra	Angola	Argentina	Armenia	Aruba	Australia	...
-------------	---------	---------	---------	--------	-----------	---------	-------	-----------	-----



DIMENSION FILTERS

Hierarchy

[Collapse all](#) | [Expand all](#)

Topic

- Education
- Environment
- Economic Policy & Debt
- Financial Sector
- Health
- Infrastructure
- Labor & Social Protection
- Poverty
- Private Sector & Trade
- Public Sector

COUNTRY

(Available: 214 Selected: 1)

SERIES

(Available: 1300 Selected: 0)

Type keywords to filter

(| A | B | C | D | E | F | G | H | I | L | M | N | O | P | Q | R | S | T | U | V | W |)

[Select all](#) | [Unselect all](#) | [Invert selection](#)

(

- (%) Benefits held by 1st 20% population - All Social Insurance
- (%) Benefits held by 1st 20% population - All Social Protection
- (%) Benefits held by 1st 20% population - All Social Safety Nets
- (%) Benefits held by 1st 20% population - Unemp benefits and ALMP
- (%) Generosity of All Social Insurance
- (%) Generosity of All Social Protection
- (%) Generosity of All Social Safety Nets
- (%) Generosity of Unemp benefits and ALMP
- (%) Program participation - All Social Insurance
- (%) Program participation - All Social Protection
- (%) Program participation - All Social Safety Nets
- (%) Program participation - Unemp benefits and ALMP

Top ▲

A

- ARI treatment (% of children under 5 taken to a health provider)
- Access to electricity (% of population)
- Adjusted net enrollment rate, primary (% of primary school age children)
- Adjusted net enrollment rate, primary, female (% of primary school age children)
- Adjusted net enrollment rate, primary, male (% of primary school age children)
- Adjusted net national income (annual % growth)
- Adjusted net national income (constant 2005 US\$)
- Adjusted net national income (current US\$)
- Adjusted net savings, excluding particulate emission damage (% of GNI)
- Adjusted net savings, excluding particulate emission damage (current US\$)
- Adjusted net savings, including particulate emission damage (% of GNI)

Top ▲

TIME

(Available: 0 Selected: 0)

Data sets

214 countries



1300 indicators

DIMENSION FILTERS**Hierarchy**[Collapse all](#) | [Expand all](#) **Year****COUNTRY**

(Available:214| Selected:1)

SERIES

(Available:1300| Selected:0)

TIME

(Available:54| Selected:0)

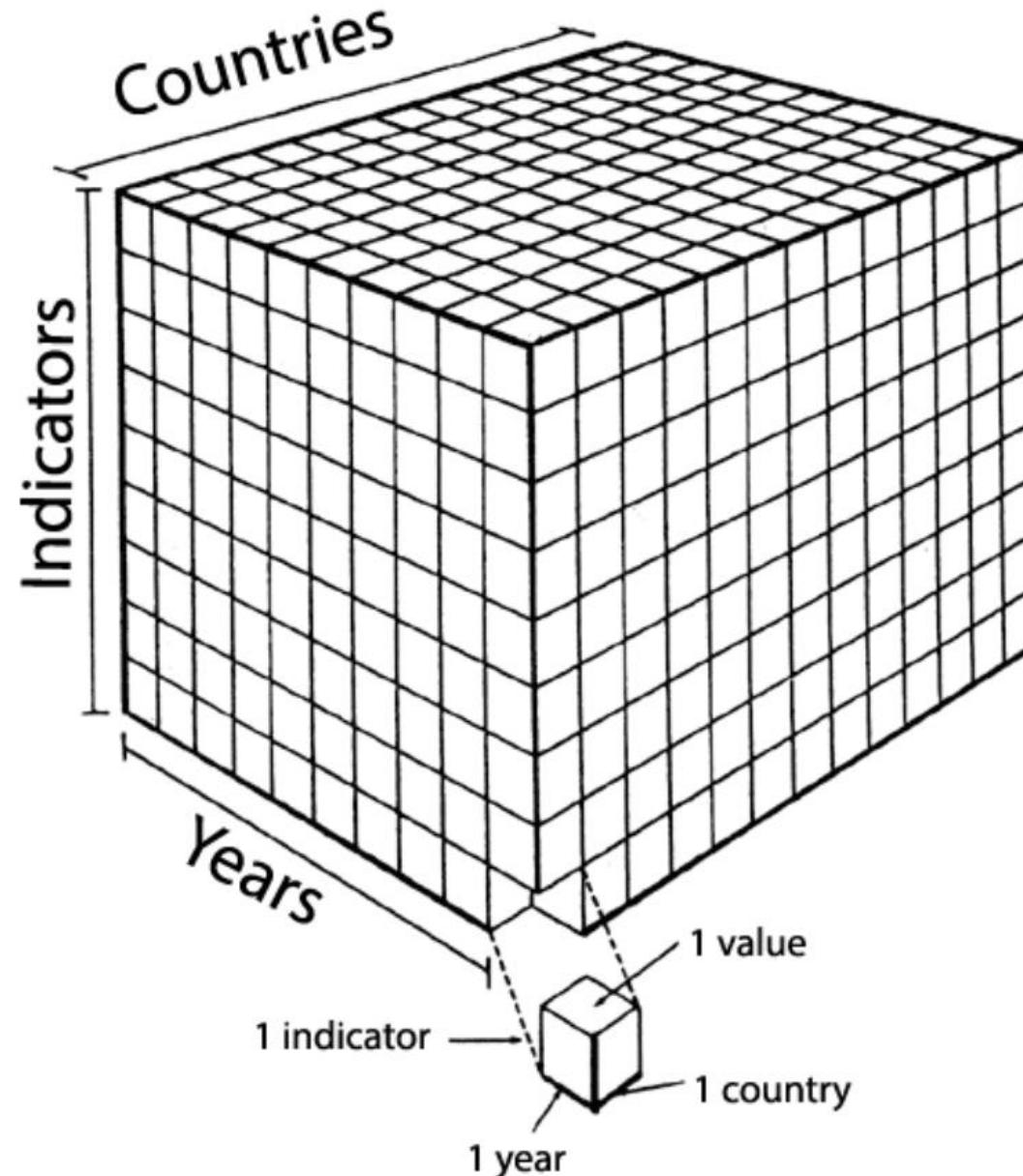
SELECT YEARS**FUNCTIONS**Availability Range: **Year 1960 - 2013**

Please drag the slider handle below to select the range

SEARCH[Select all](#) | [Unselect all](#) | [Invert selection](#) 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994

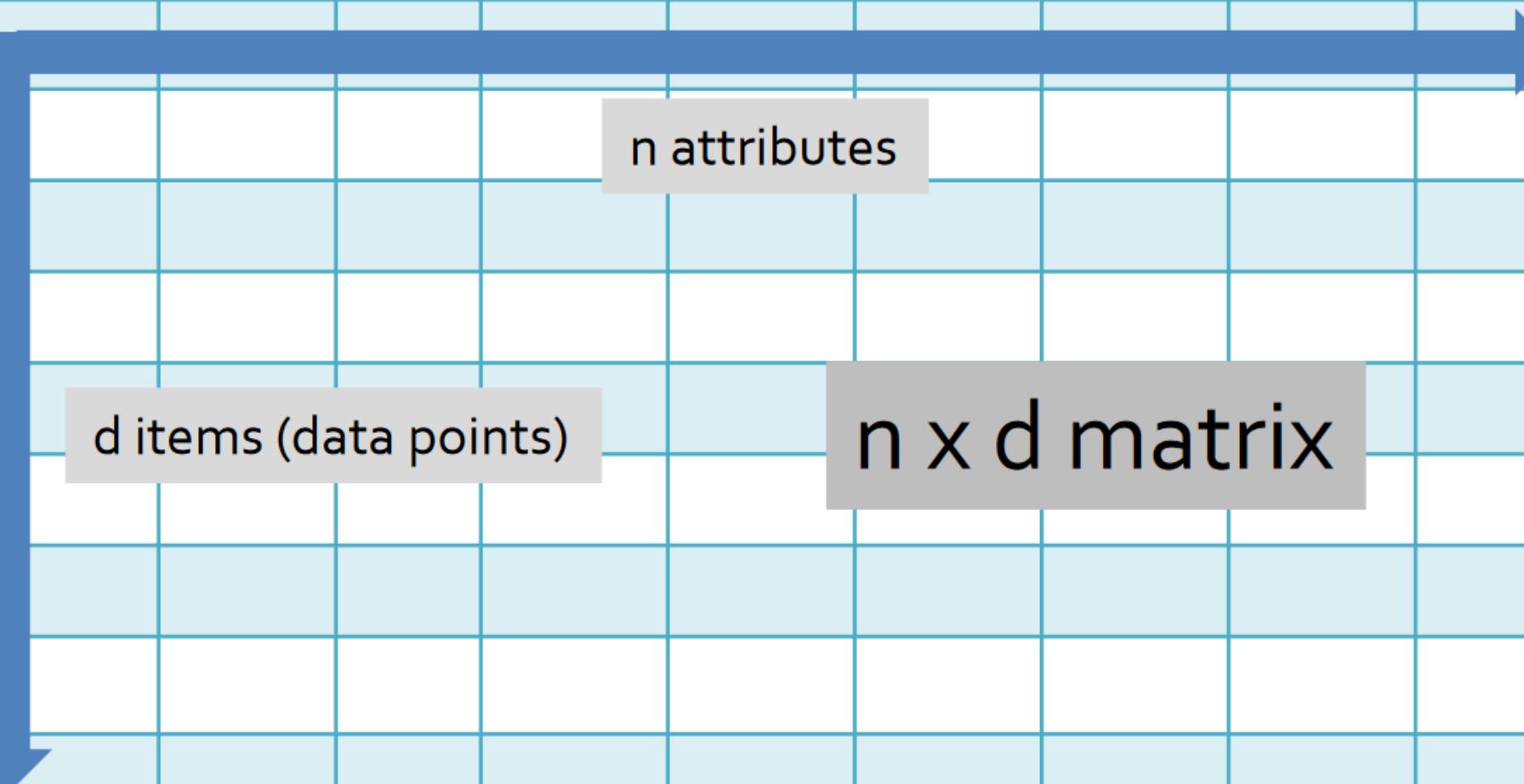
Data sets

- Data Cube



Data sets

	Afghanistan	Albania	Algeria	Andorra	Angola	Argentina	Armenia	Aruba	Australia	...
Benefits										
Generosity										
Participation										
Savings										
Interest										
Bird species										
Birth Rate										
C02 Emissions										
...										

A diagram illustrating a data matrix. A vertical blue arrow on the left points downwards, labeled "d items (data points)". A horizontal blue arrow at the top points to the right, labeled "n attributes". A large grey box on the right is labeled "n x d matrix". The matrix itself is a grid of light blue cells, with row labels on the left and column labels at the top.

Outline

- *Data sets*
- **Layout exploration**
- Layouts for multi-dimensional data

Layout exploration

- Generate visualization
 - Find layout (arrange tabular data)
 - Choose visual encoding and marks

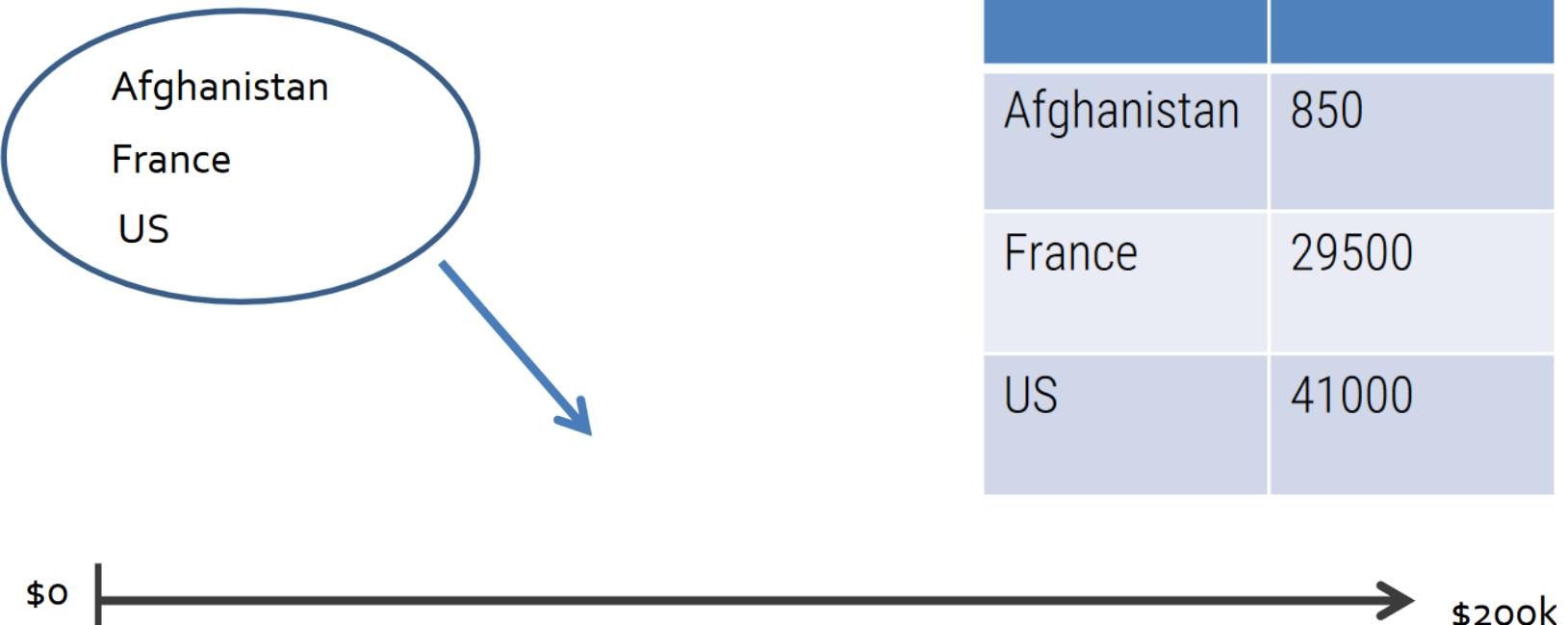
Layout exploration

- First task: find layout
 - That provides an understanding of the data
- Let's start with two attributes
 - Country & income per person, e.g.:

Country	Income per person
Afghanistan	850
France	29500
US	41000

Layout exploration

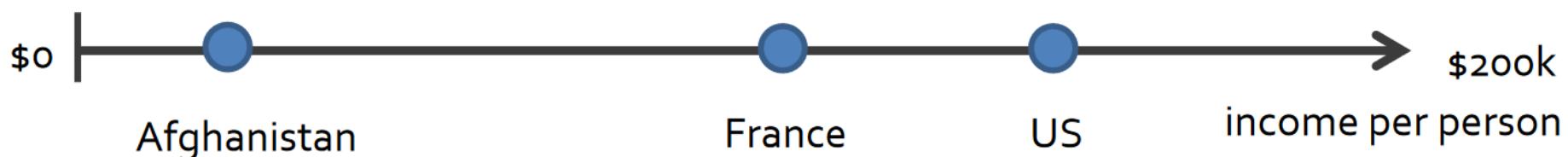
- Find a layout



Layout exploration

- Find a layout
 - E.g. position + circle

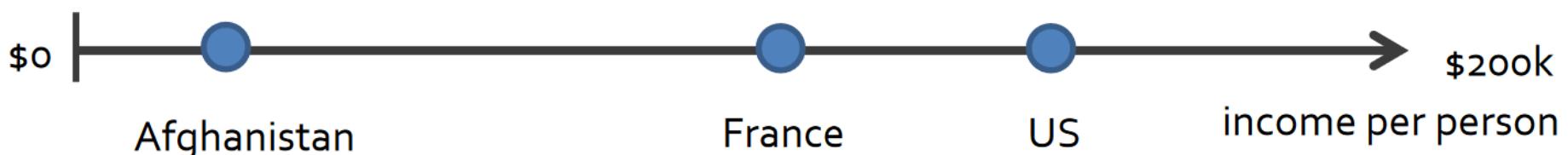
Country	Income per person
Afghanistan	850
France	29500
US	41000



Layout exploration

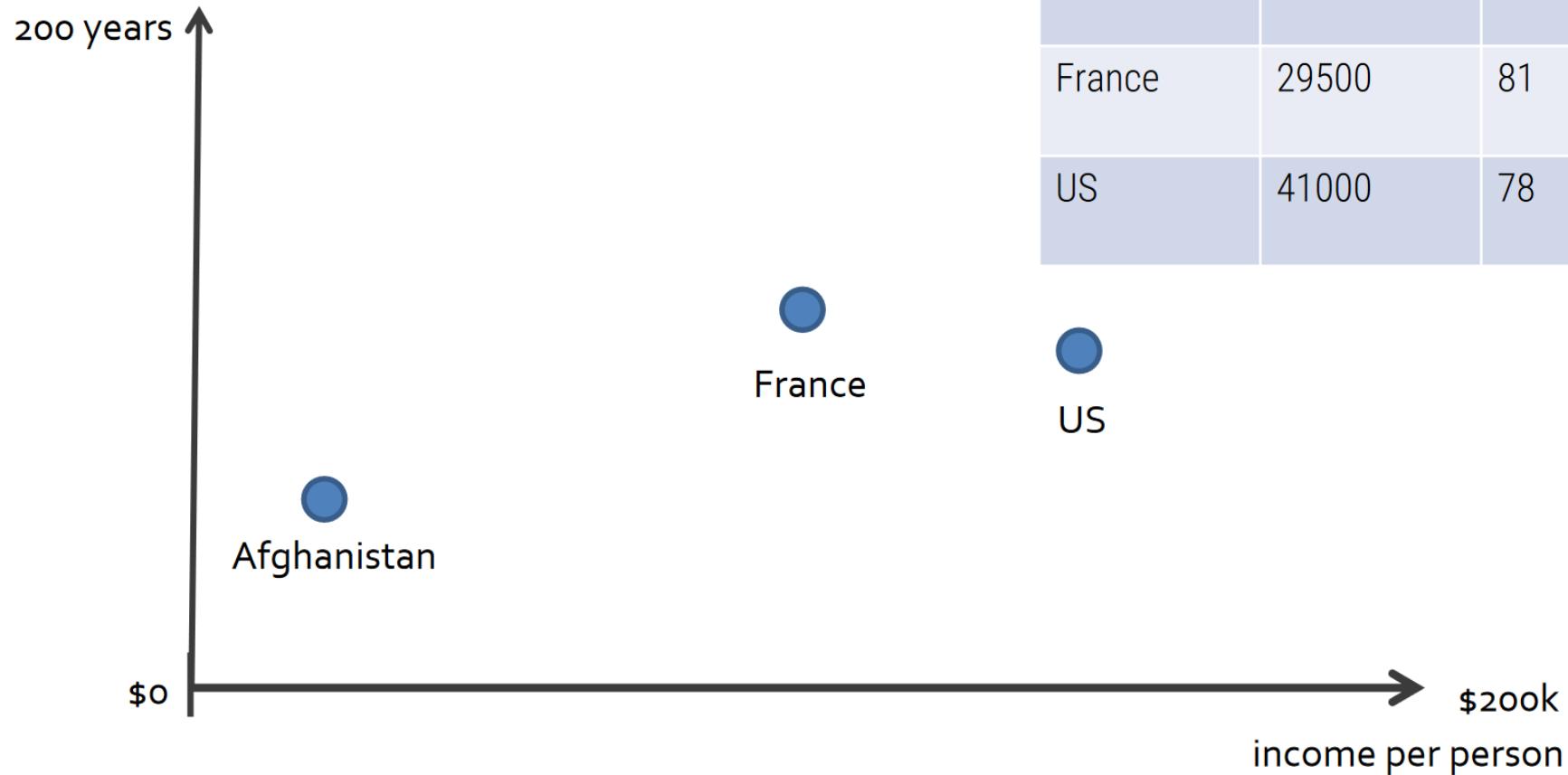
- Find a layout
 - How to extend to 3 attributes?

Country	Income per person	Life expectancy
Afghanistan	850	57
France	29500	81
US	41000	78



Layout exploration

- Extend the layout to 3 attributes

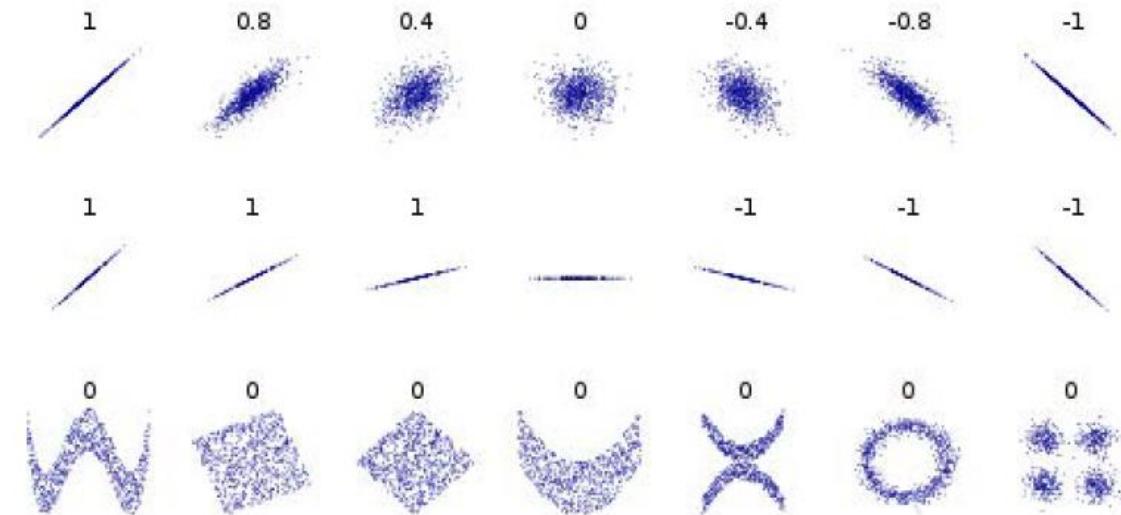
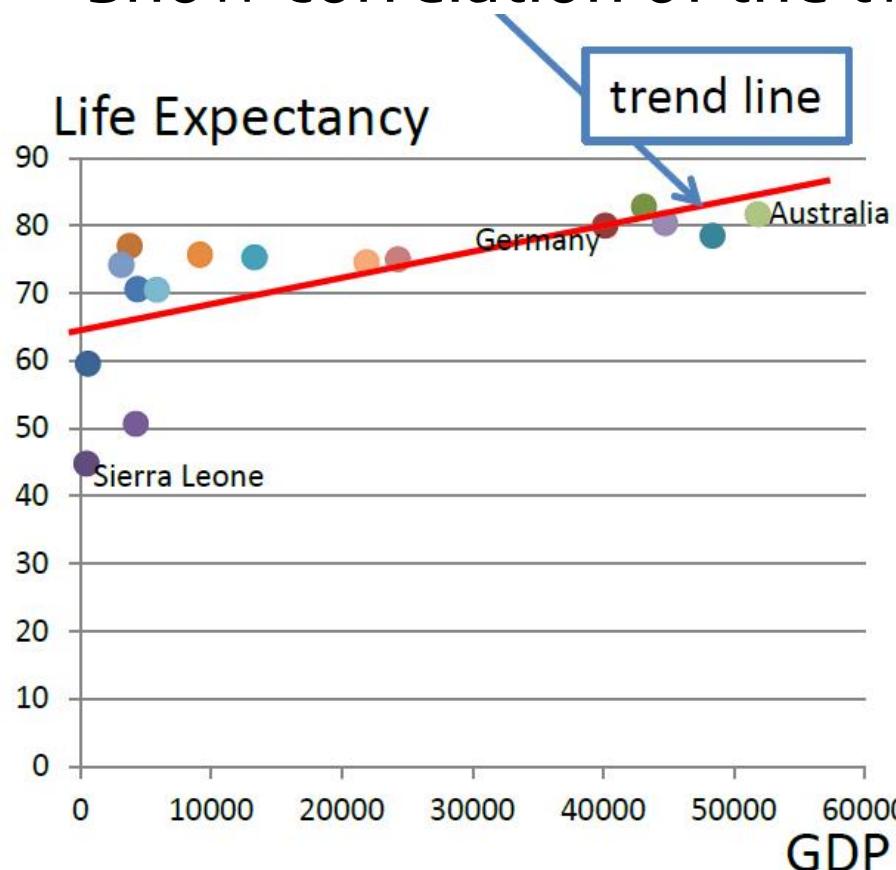


Layout exploration

- Scatterplots
 - Two quantitative values
 - Horizontal and vertical spatial dimensions
 - Mark type = point

Layout exploration

- Scatterplots
 - Show correlation of the two attributes



Correlation patterns. Numbers represent the Pearson correlation coefficient.

Layout exploration

- Adding additional attributes to scatterplots
 - Color?
 - Point sizes?
 - Third dimension?
- Scatterplots with sized marks are often called bubble charts/plots

Layout exploration

- Cautions with extra attributes
 - Not all parameters allow same number of discriminable values
 - Might lead to clutter (e.g. if too many values)

Life expectancy, years [?](#)Income per person, GDP/capita in \$/year adjusted for inflation & prices [?](#)Color [World Regions](#) [?](#)Select Search... [?](#)

- Afghanistan
- Albania
- Algeria
- Andorra
- Angola
- Antigua and Barbuda
- Argentina
- Armenia
- Aruba
- Australia
- Austria
- Azerbaijan
- Bahamas
- Bahrain
- Bangladesh
- Barbados
- Belarus
- Belgium
- Belize
- Benin

Size [Population](#) [?](#)Zoom [100%](#) [OPTIONS](#) [EXPAND](#) [PRESENT](#)

Outline

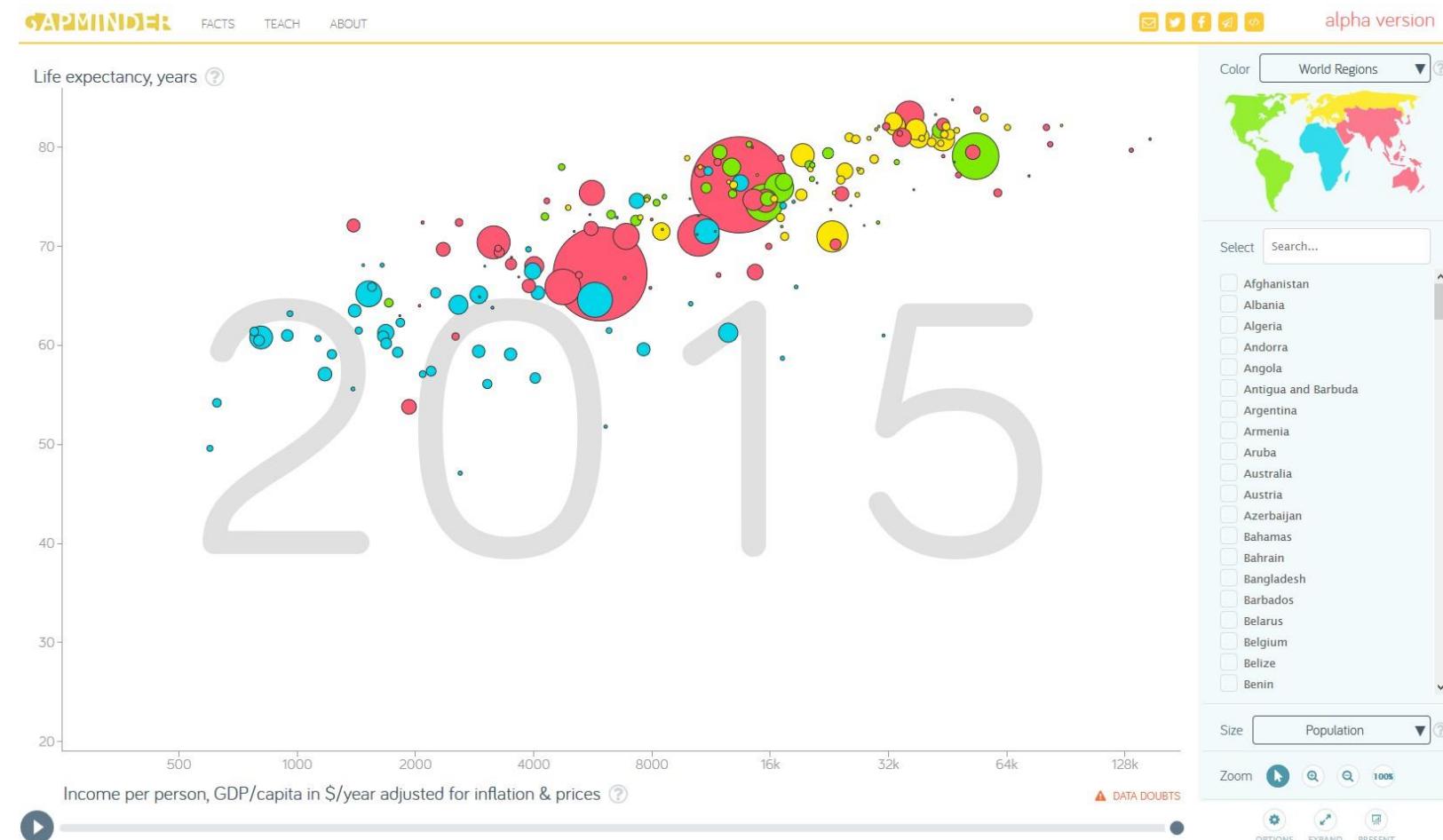
- *Data sets*
- *Layout exploration*
- **Layouts for multi-dimensional data**

Layouts for multi-dimensional data

- Visualization is a design process
 - Design
 - Evaluate
 - Improve
 - Iterate

Layouts for multi-dimensional data

- Scatterplots: Tasks
 - Find trends
 - Find outliers
 - Show distribution
 - Show correlation
 - Locate clusters

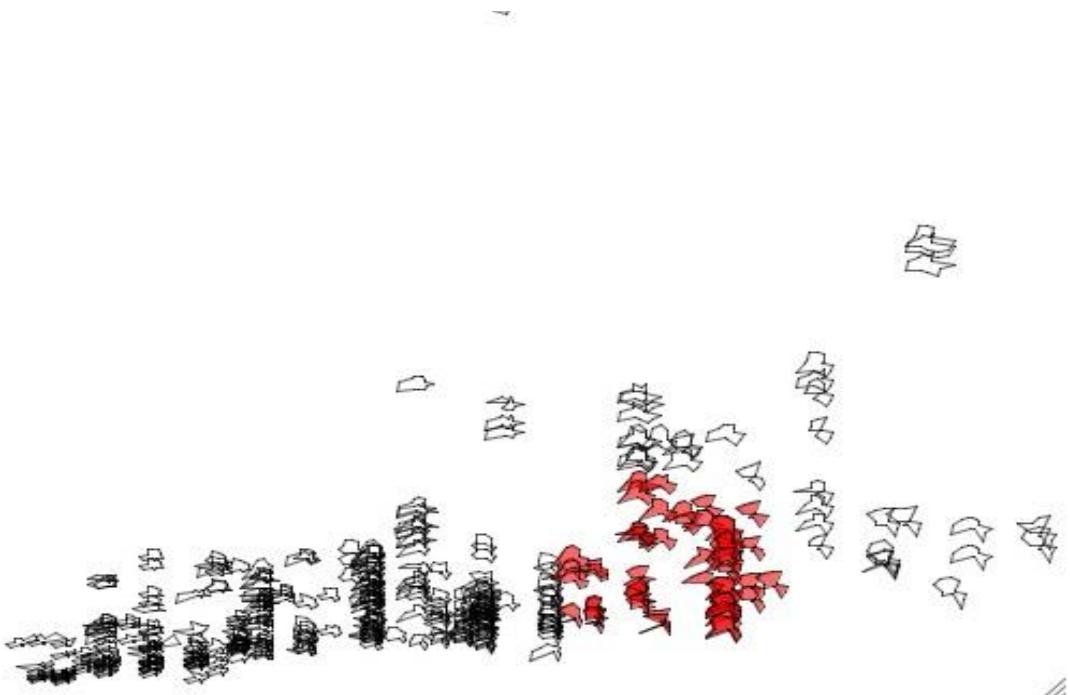


Layouts for multi-dimensional data

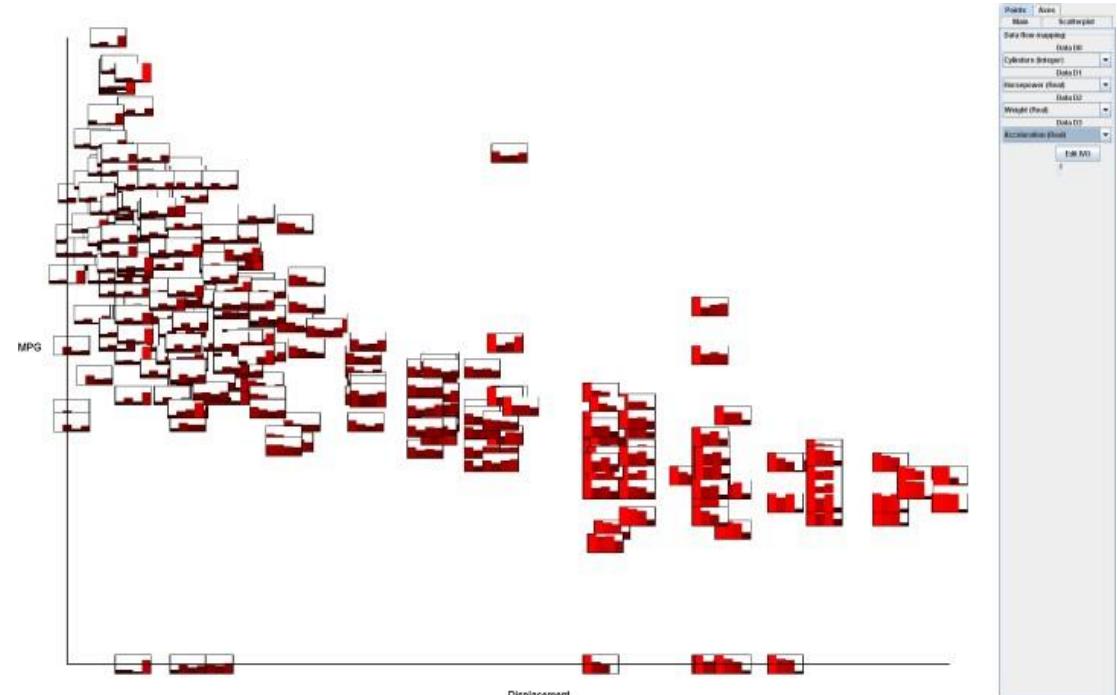
- How many items are reasonable to put on a scatterplot?
 - Hundreds of items
 - Though some extra technique may be required if cluttering appears

Layouts for multi-dimensional data

- Marks can be replaced with glyphs: composite object, internal structure from multiple marks



<http://rosuda.org/software/Gauguin/gauguin.html>



<https://engineering.purdue.edu/~elm/projects/gpuvis.html>

Layouts for multi-dimensional data

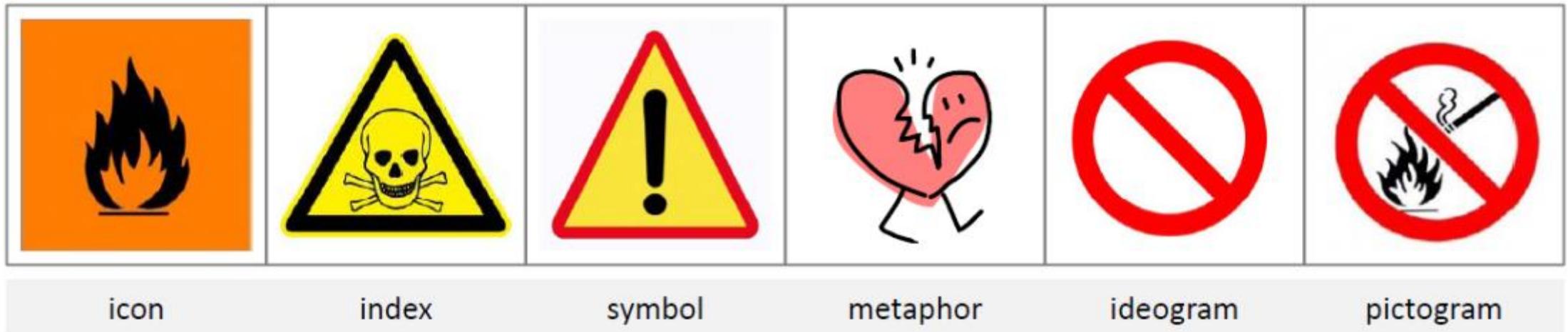
- **Glyphs**
 - Small visual object that can be used independently and constructively to depict attributes of a data record or the composition of a set of data records
 - Typically free from reference structures
 - Grid lines, axes labels...

Layouts for multi-dimensional data

- **Glyphs**
 - Can be placed spatially independently from others
 - Though in some cases are spatially connected to convey topological relationships between data records or geometric discontinuity
 - Type of visual signs but that can make use of other types of signs such as icons

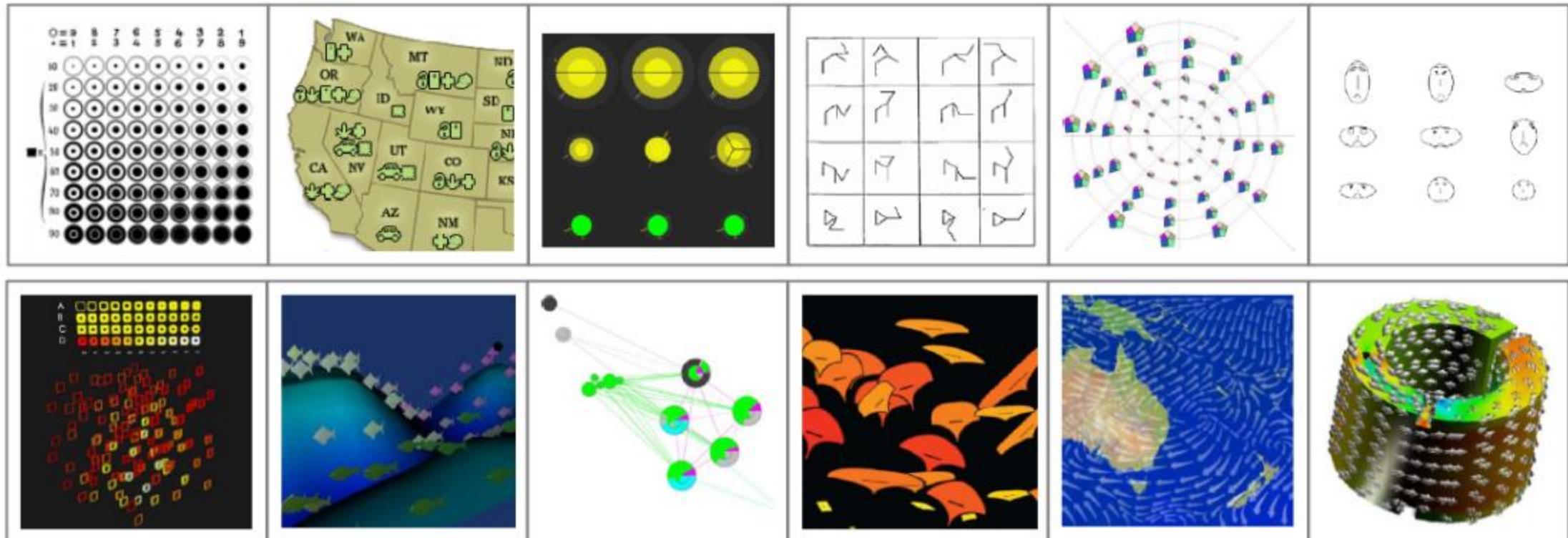
Layouts for multi-dimensional data

- Glyphs vs symbols
 - Symbols (and other): **not linked with data records**



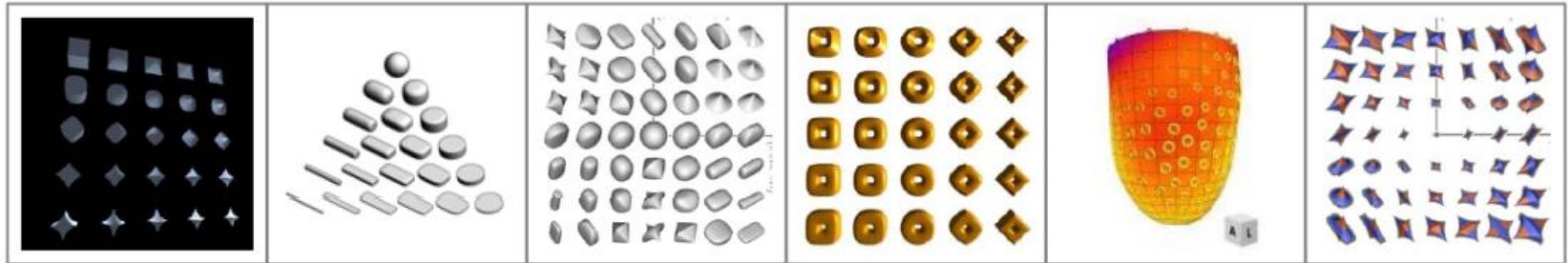
Layouts for multi-dimensional data

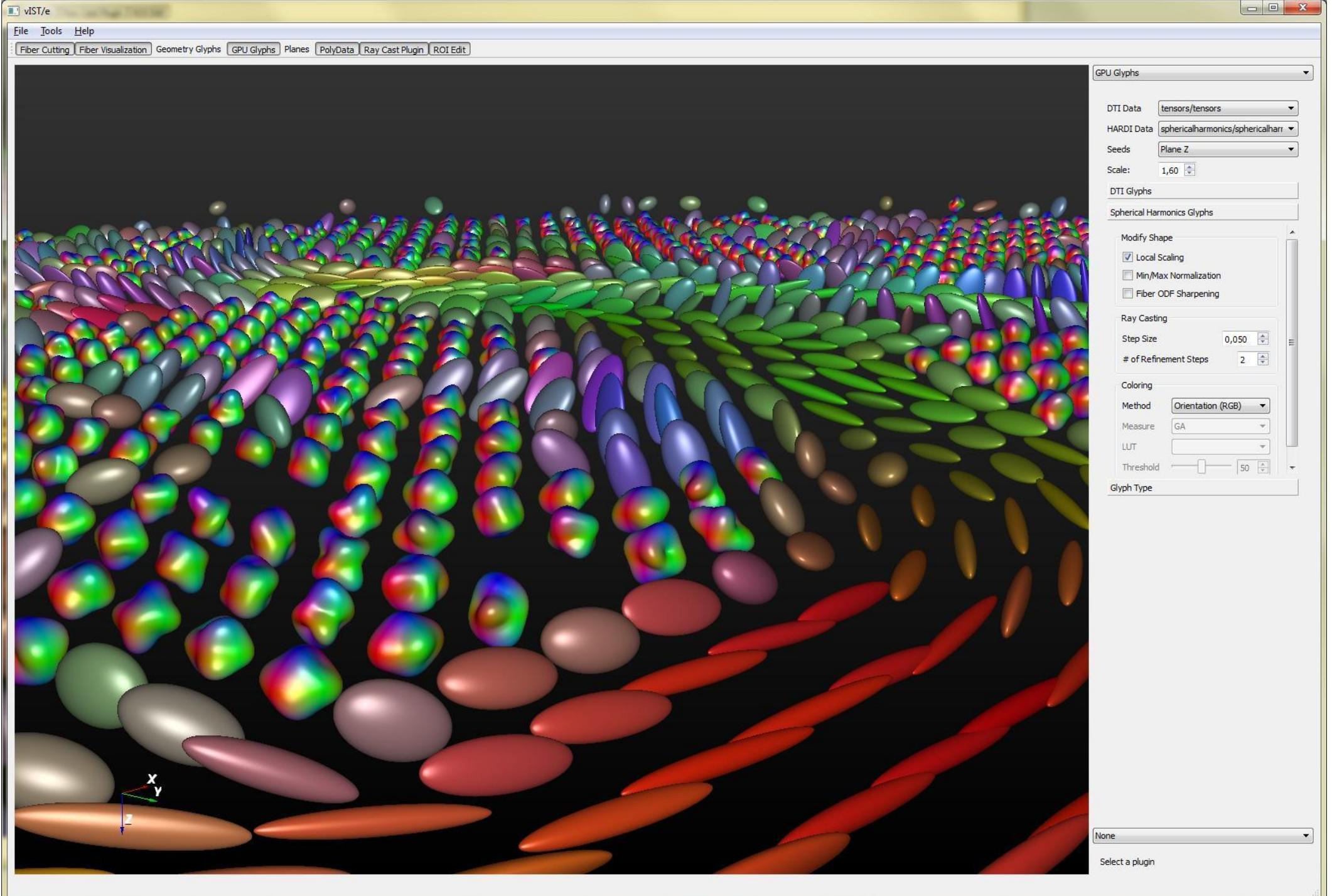
- Glyphs carry (encode) information on data records



Layouts for multi-dimensional data

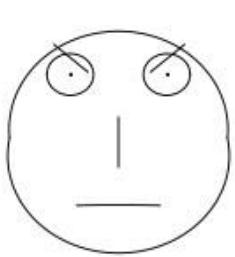
- Glyphs carry (encode) information on data records



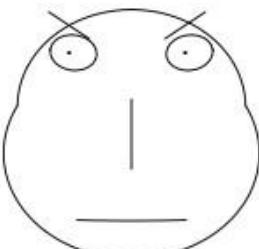


Layouts for multi-dimensional data

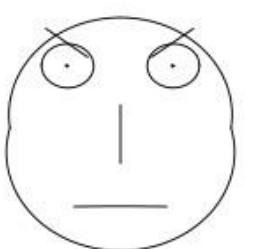
- Glyphs: Chernoff faces [Chernoff73]



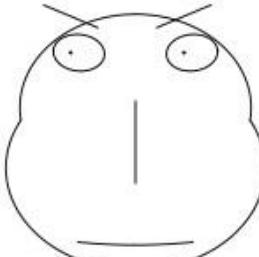
AARONSON,L.H.



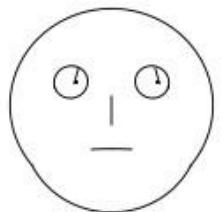
ALEXANDER,J.M.



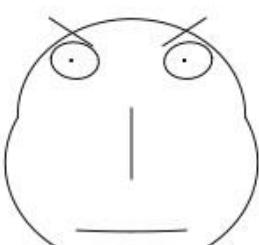
ARMENTANO,A.J.



BERDON,R.I.



BRACKEN,J.J.



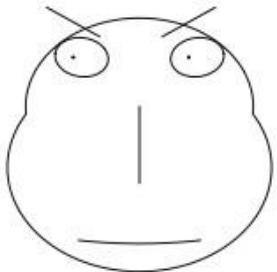
BURNS,E.B.



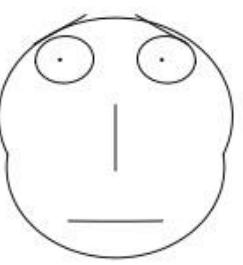
CALLAHAN,R.J.



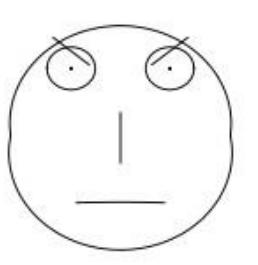
COHEN,S.S.



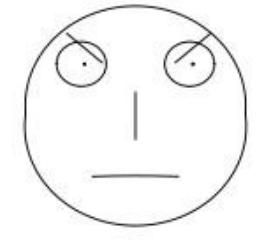
DALY,J.J.



DANNEHY,J.F.



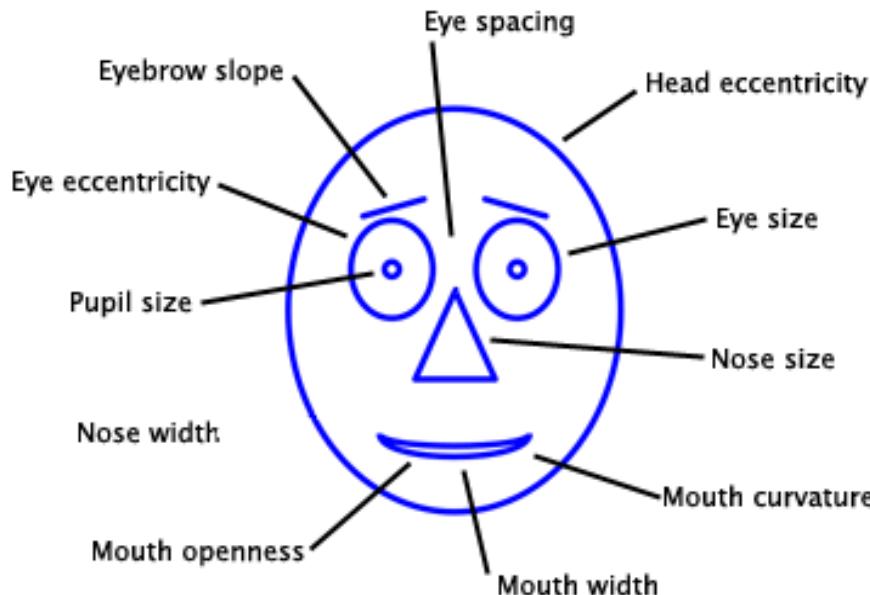
DEAN,H.H.



DEVITA,H.J.

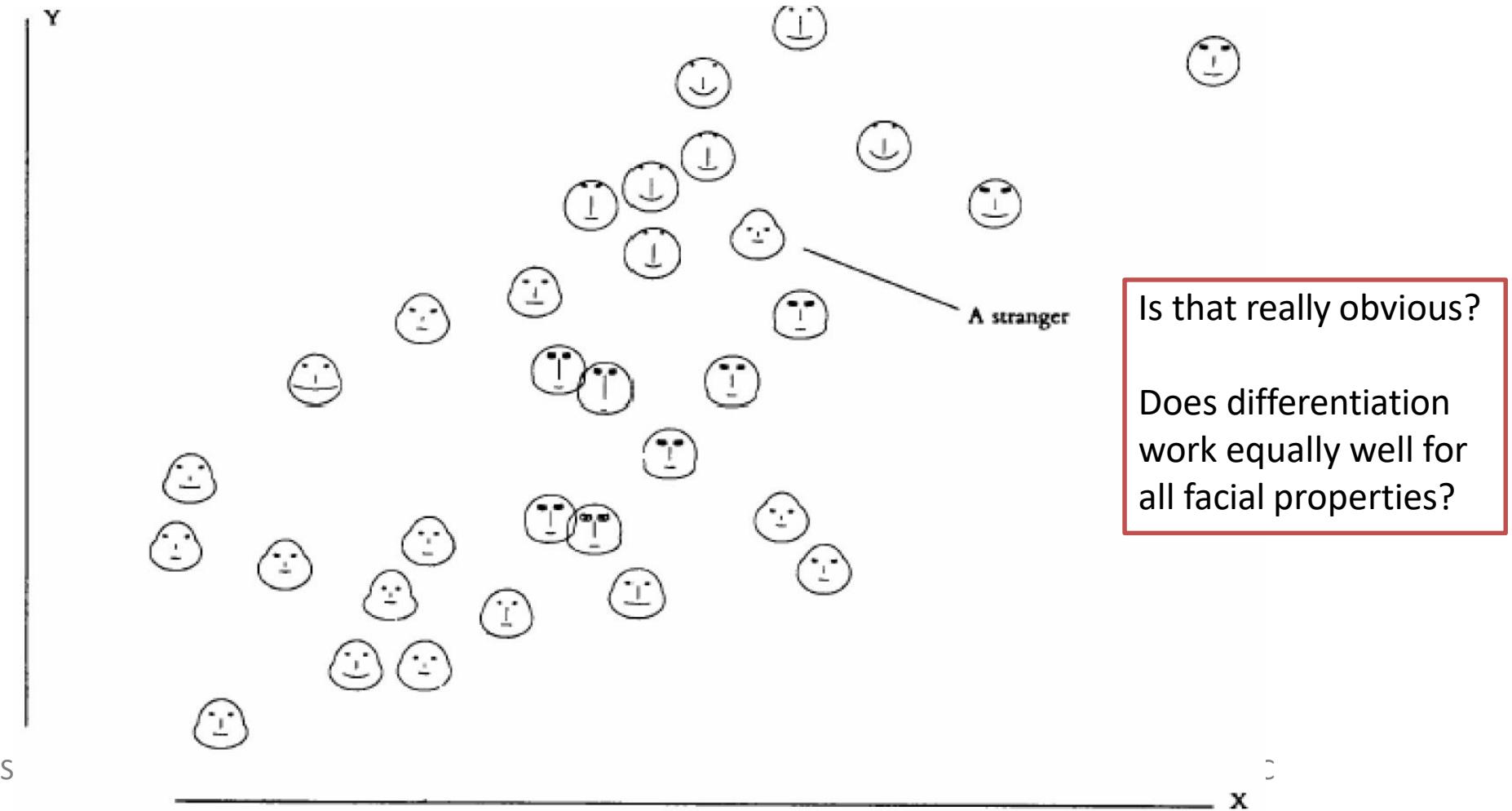
Layouts for multi-dimensional data

- Glyphs: Chernoff faces
 - Up to 18 attributes can be mapped to face properties
 - Humans are familiar with interpreting faces
 - High variance in interpretation of the same face



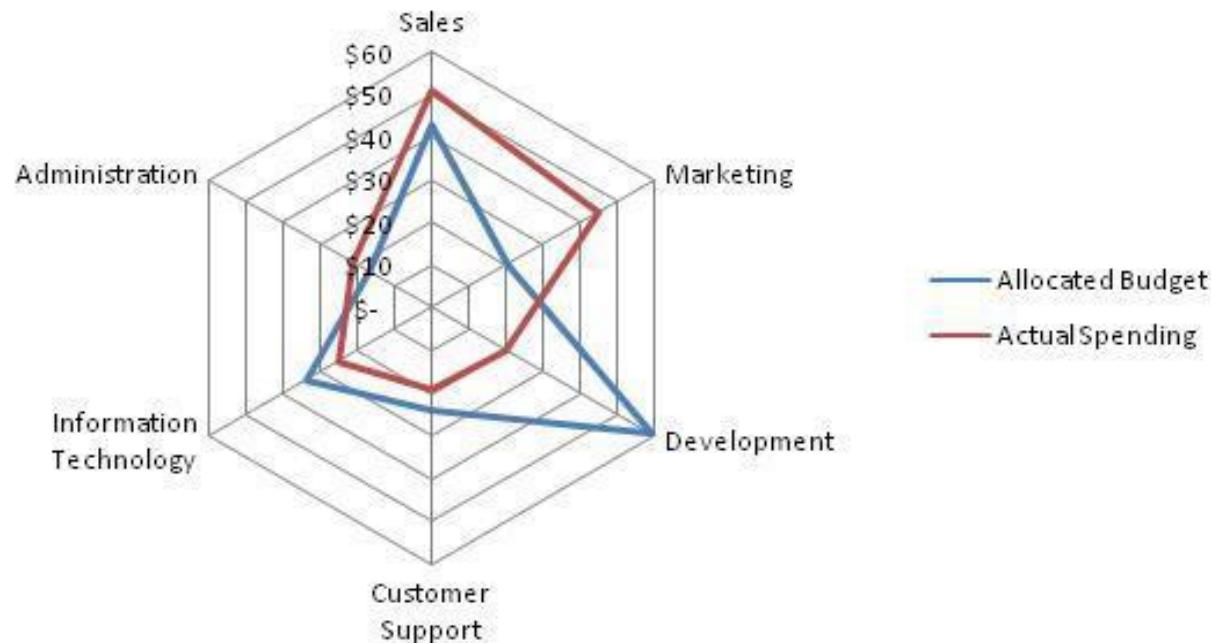
Layouts for multi-dimensional data

- Glyphs: Chernoff faces



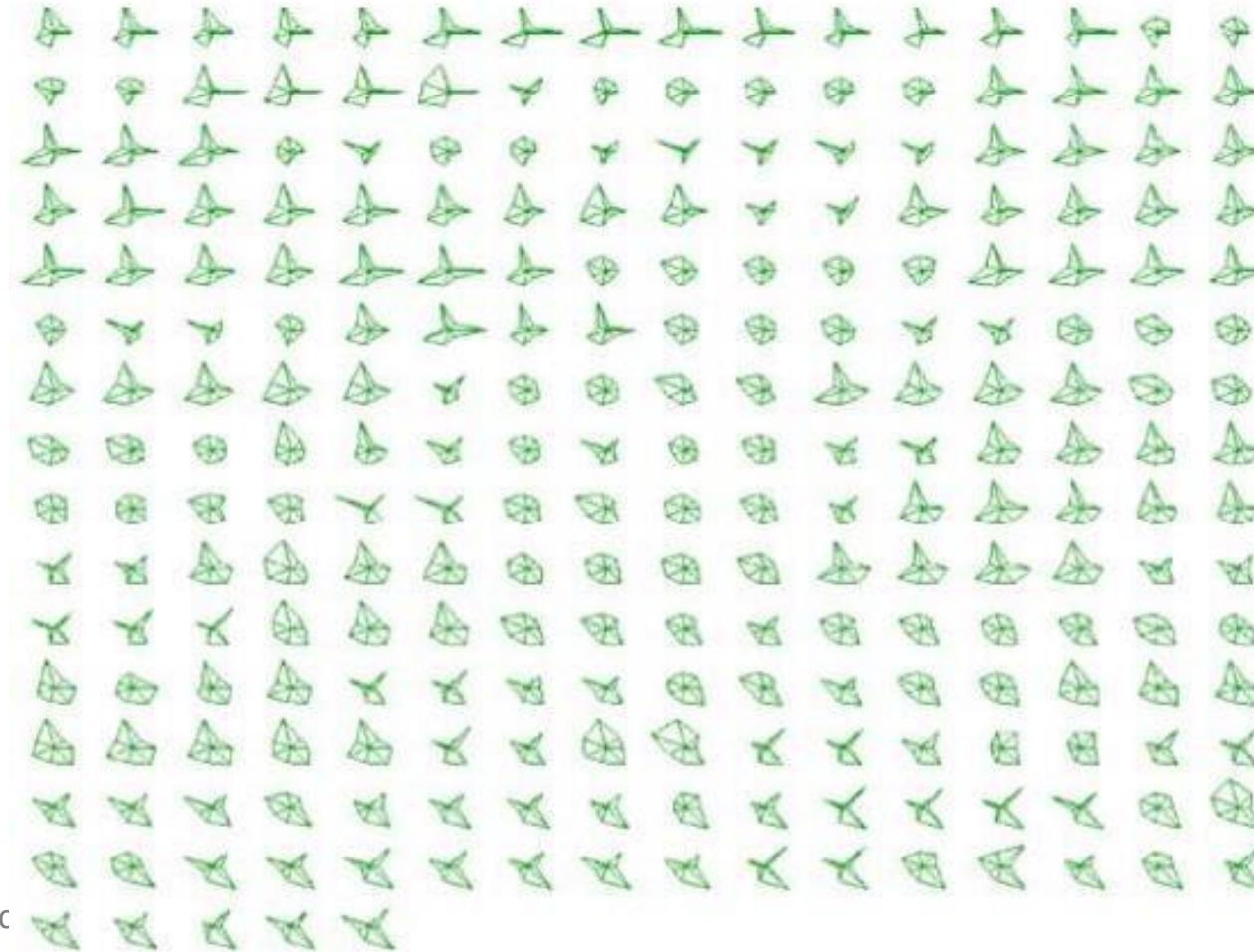
Layouts for multi-dimensional data

- Glyphs: Star glyphs
 - Lay out dimension in radial fashion
 - Draw each point as a ring



Layouts for multi-dimensional data

- Glyphs: Star glyphs

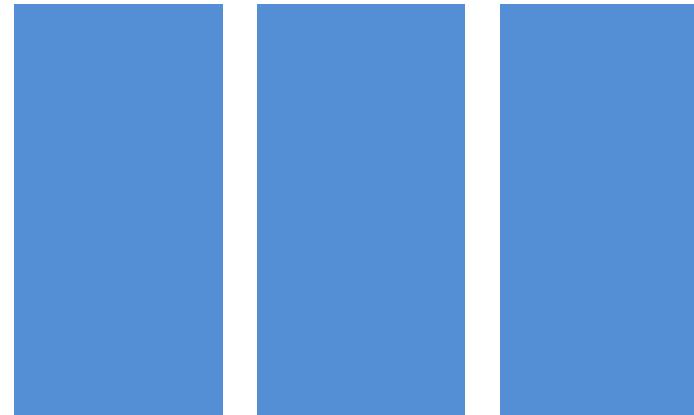


Layouts for multi-dimensional data

- Showing categorical regions
 - Solving tasks such as: Separate, Order, Align
 - Spatial position is an ordered magnitude visual channel
 - Categorical attributes are unordered identities (no magnitude)
 - → Cannot be encoded with spatial position
 - BUT: can be expressed with a spatial region

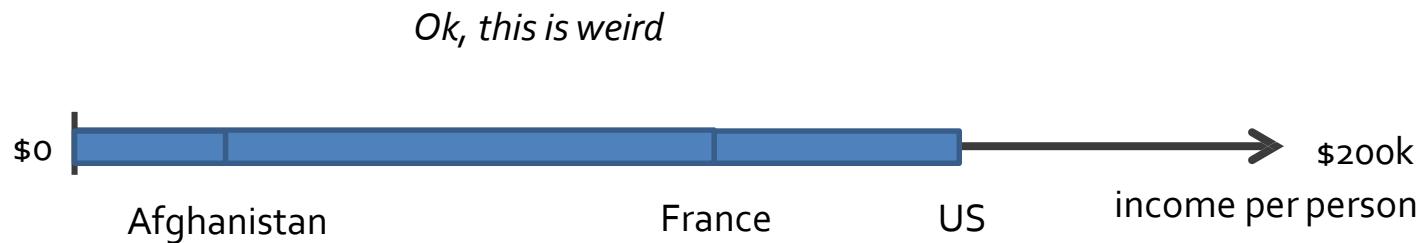
Layouts for multi-dimensional data

- Showing categorical regions
 - Contiguous bounded areas
 - Distinct from one another
 - Need to be separated, ordered, and aligned



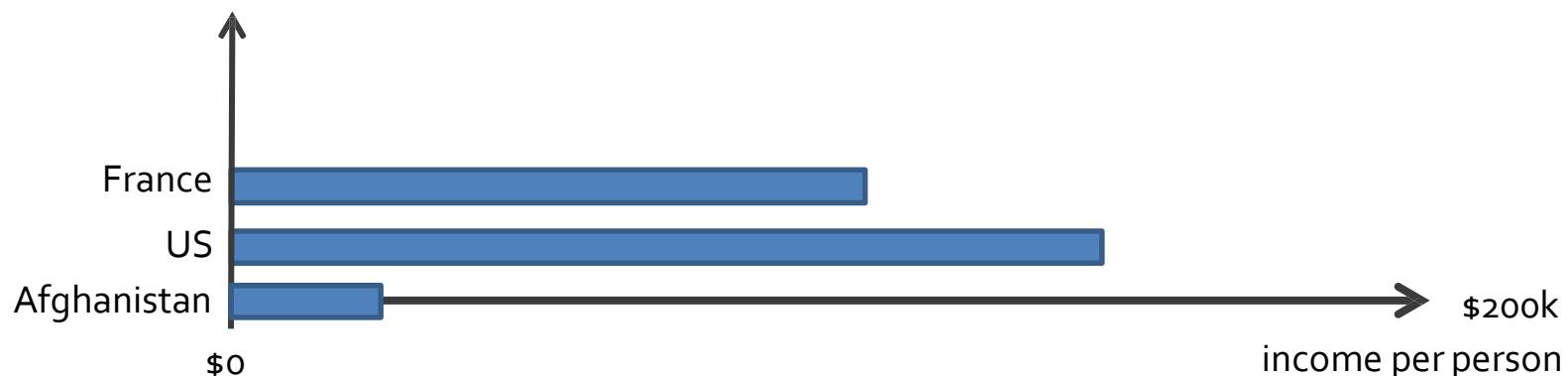
Layouts for multi-dimensional data

- List alignment
 - Need to separate into regions by key
 - E.g. length + rectangle



Layouts for multi-dimensional data

- List alignment
 - Need to separate into regions by key
 - E.g. length + rectangle
 - We can do better, e.g. aligning regions of key categorical values along one axis in a common frame

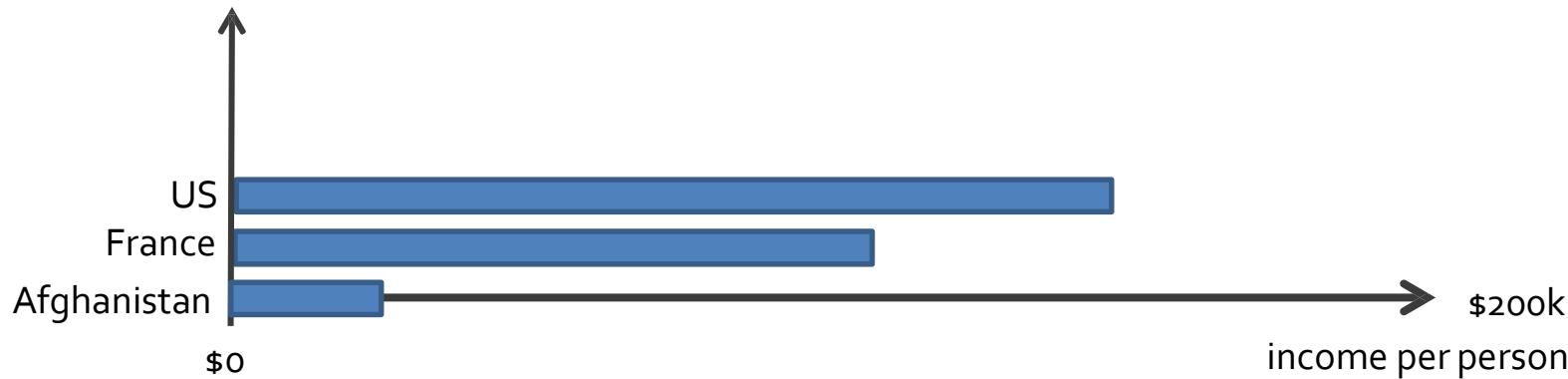


Layouts for multi-dimensional data

- Ordering
 - To facilitate search, data must be ordered
 - By some attribute
- LATCH:
 - Location
 - Alphabet
 - Time
 - Category
 - Hierarchy

Layouts for multi-dimensional data

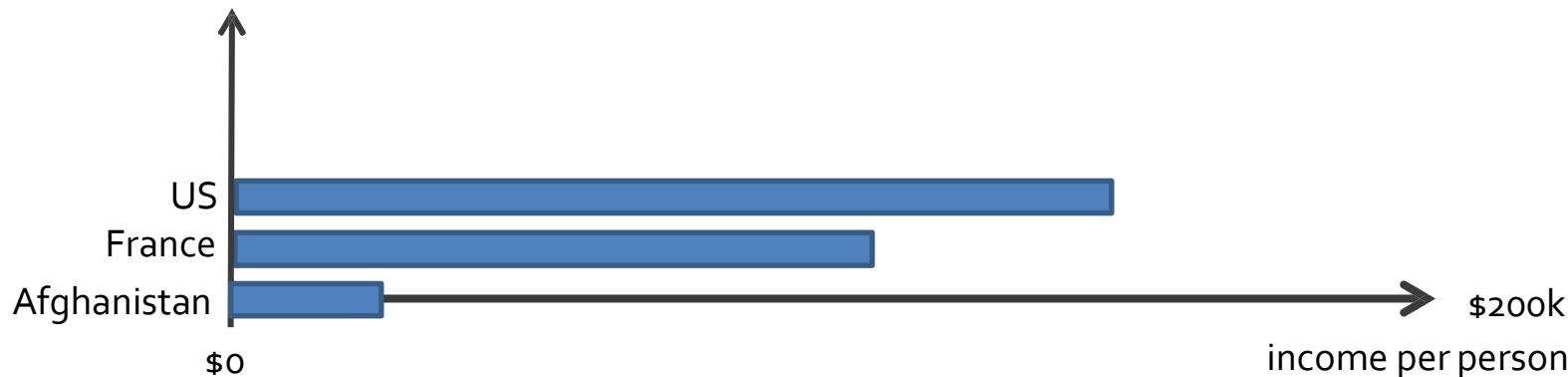
- Ordering
 - Let's use quantity



- Advantages?

Layouts for multi-dimensional data

- Ordering
 - Let's use quantity
 - Good if the quantity sorting matters!!!
 - Difficult to search for a country

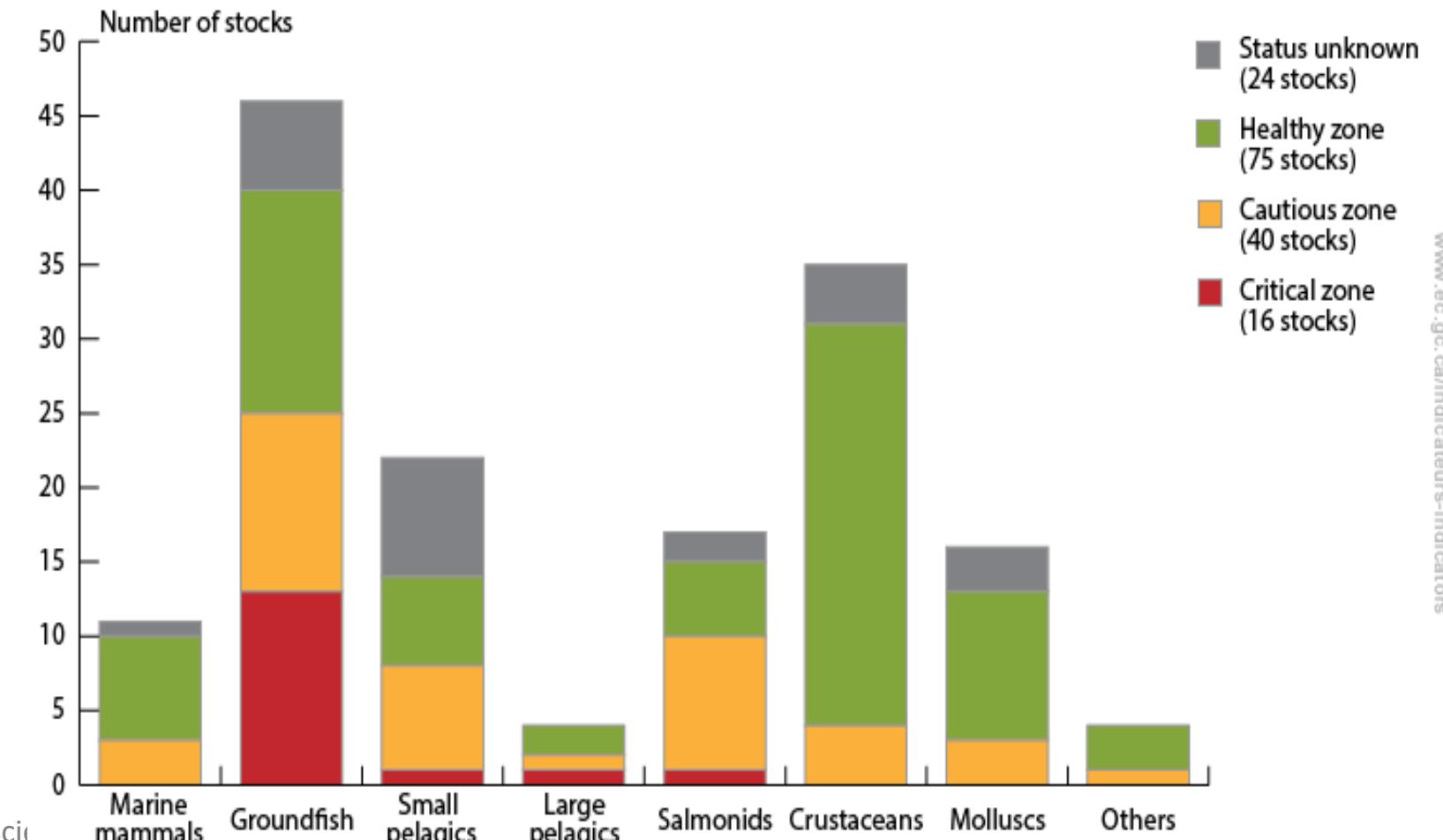


Layouts for multi-dimensional data

- Increase the number of dimensions in bar charts? → Stacked bar charts
 - Each bar is a composite glyph
 - Each bar part encodes a value
 - Composite glyphs arranged as a list according to primary key
 - Color used to distinguish
 - Secondary key
 - Typically used for absolute values (use a normalized stacked bar for proportions)

Layouts for multi-dimensional data

- Stacked bar charts

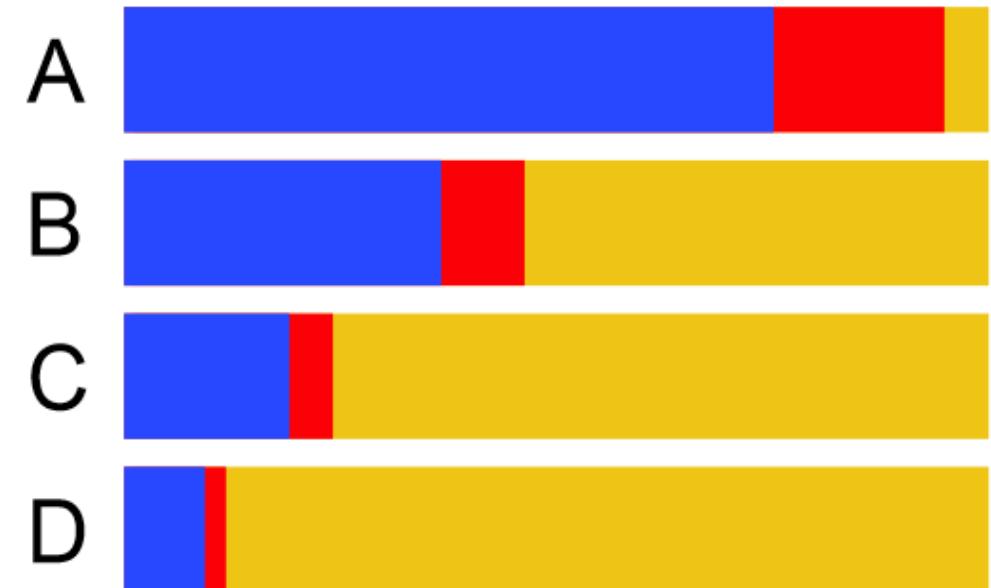
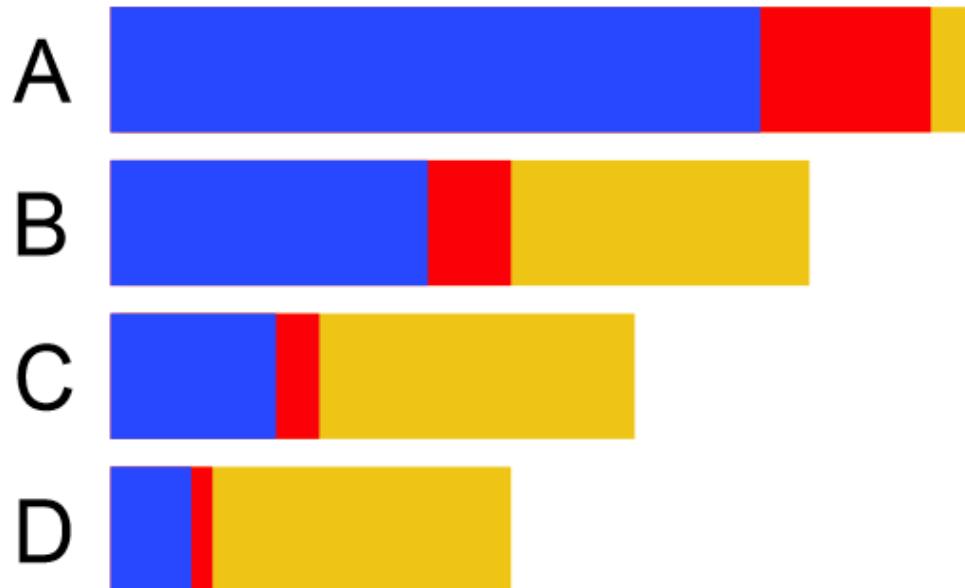


Layouts for multi-dimensional data

- Stacked bar charts
 - Can compare totals and lowest level well
 - Upper levels of secondary key require comparison against non-aligned scale

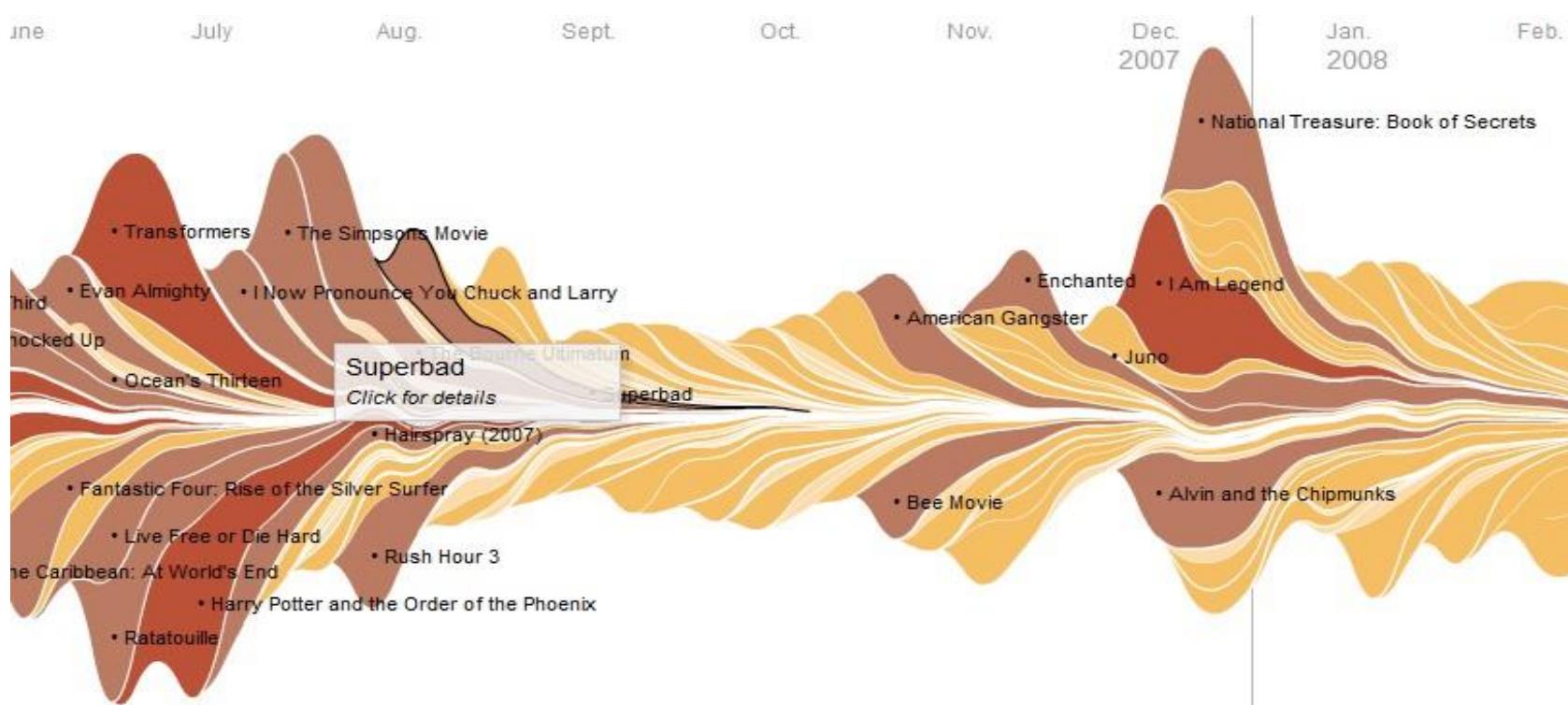
Layouts for multi-dimensional data

- Stacked bars vs normalized stacked bars
 - Can compare lower and upper levels
 - Useful when we have normalized data (e.g. percentages)



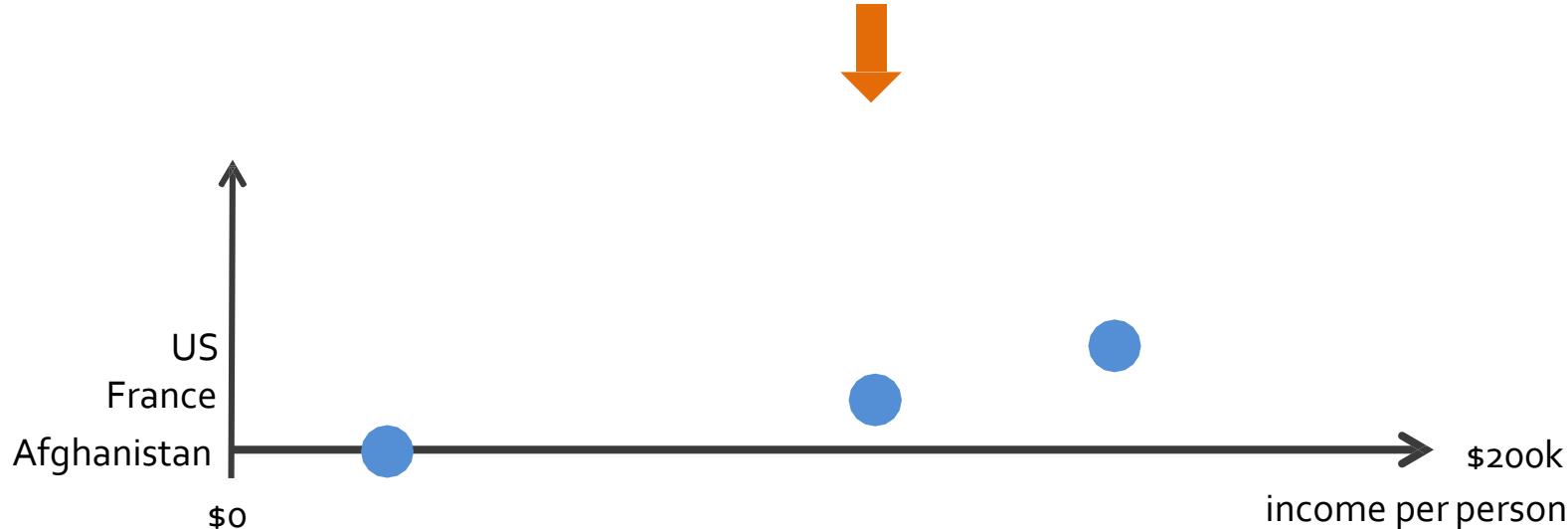
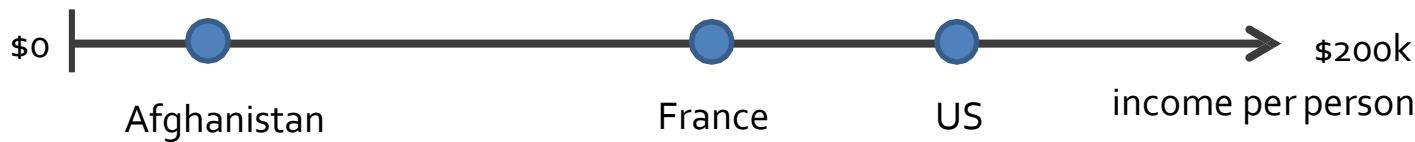
Layouts for multi-dimensional data

- Streamgraphs



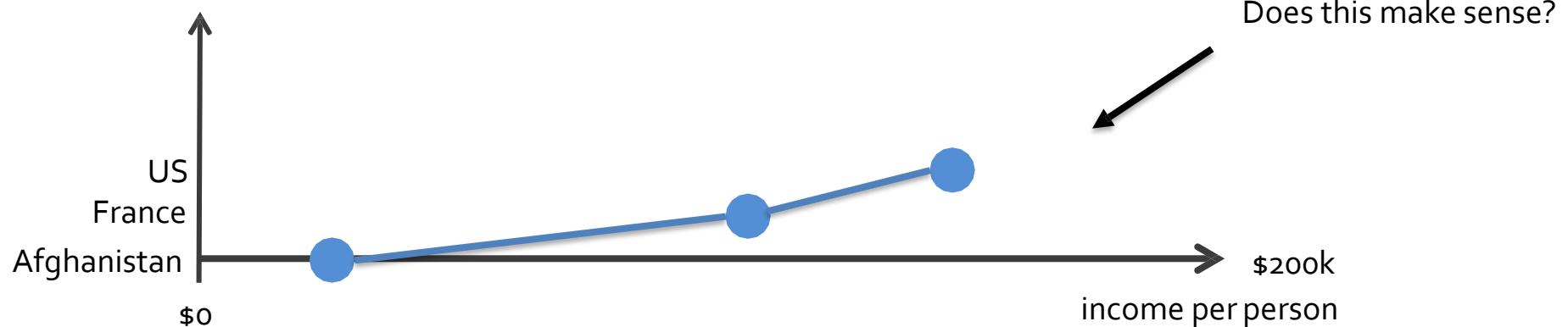
Layouts for multi-dimensional data

- Dot chart/plot



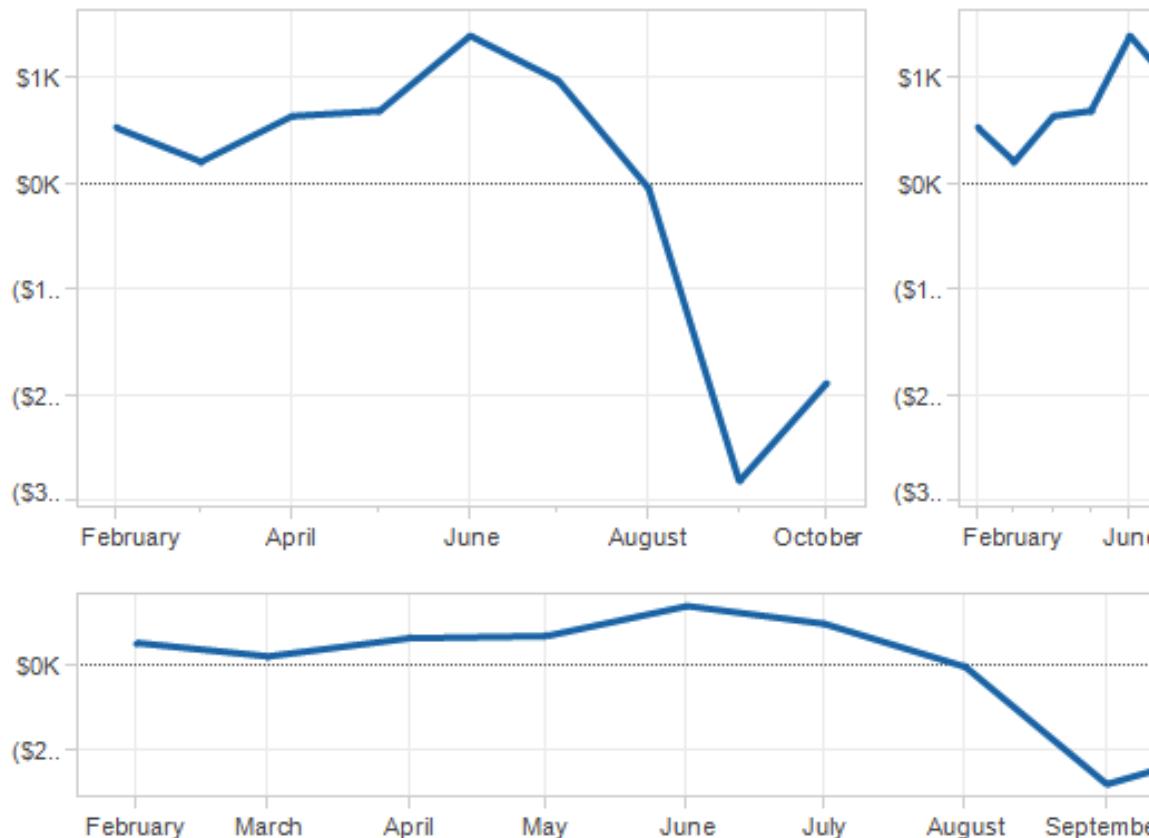
Layouts for multi-dimensional data

- Line chart
 - Augment with connection marks
 - Emphasize the ordering and show trends
 - Should not be used with categorical keys



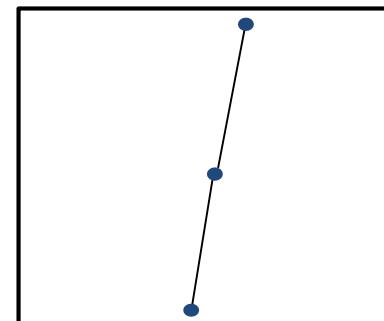
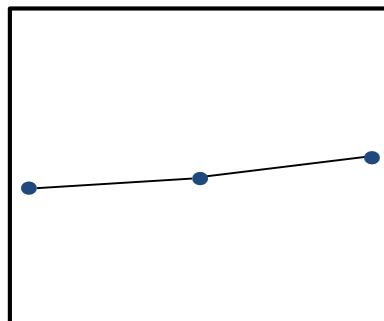
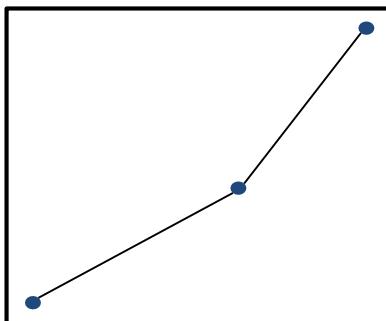
Layouts for multi-dimensional data

- Line chart: Aspect Ratio Selection



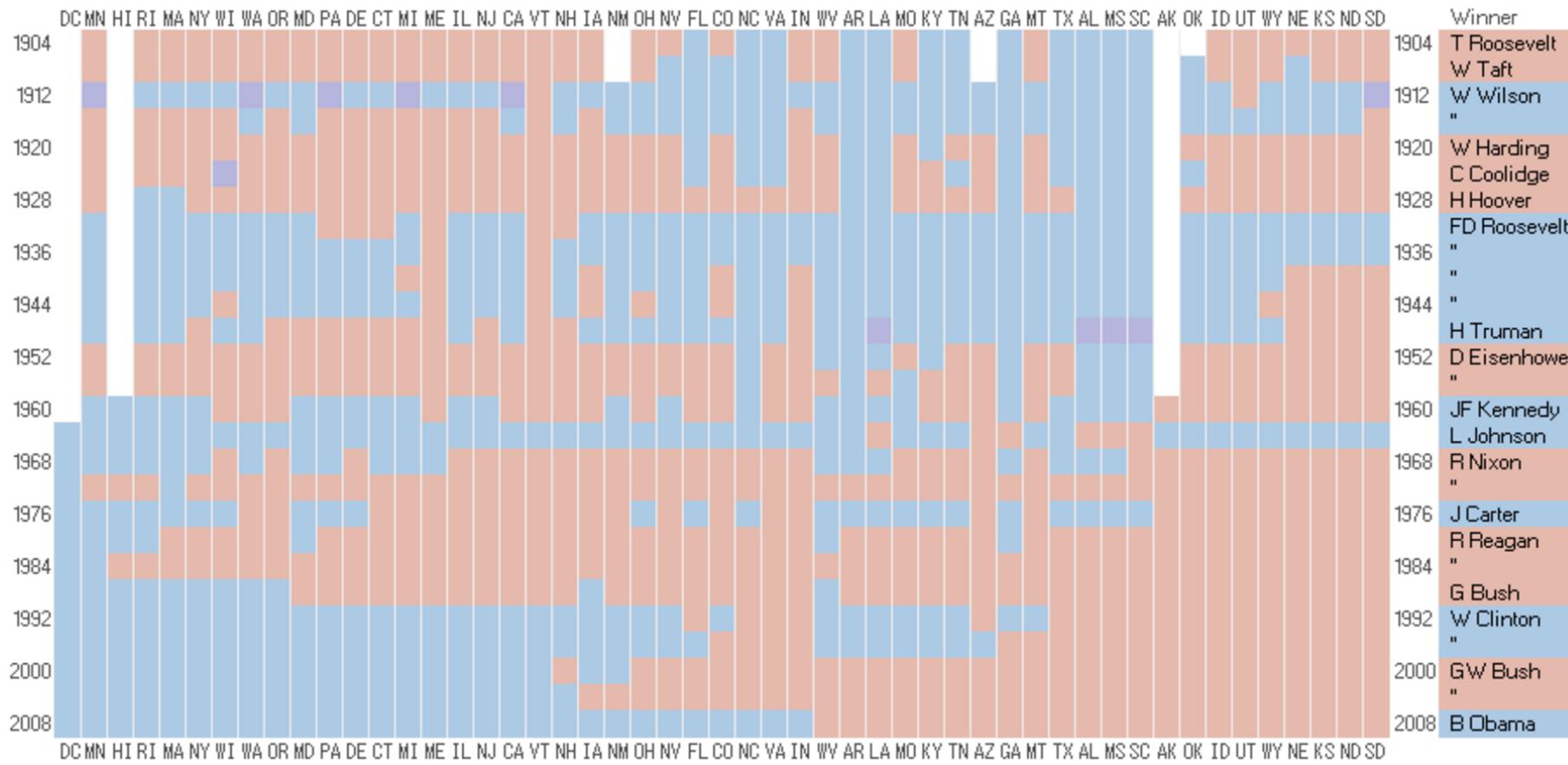
Layouts for multi-dimensional data

- Line chart: banking to 45°
 - To facilitate perception of trends, maximize the discriminability of line segment orientations
 - Two segments are maximally discriminable when their average absolute angle is 45°
 - Can optimize the aspect ratio to bank to 45°



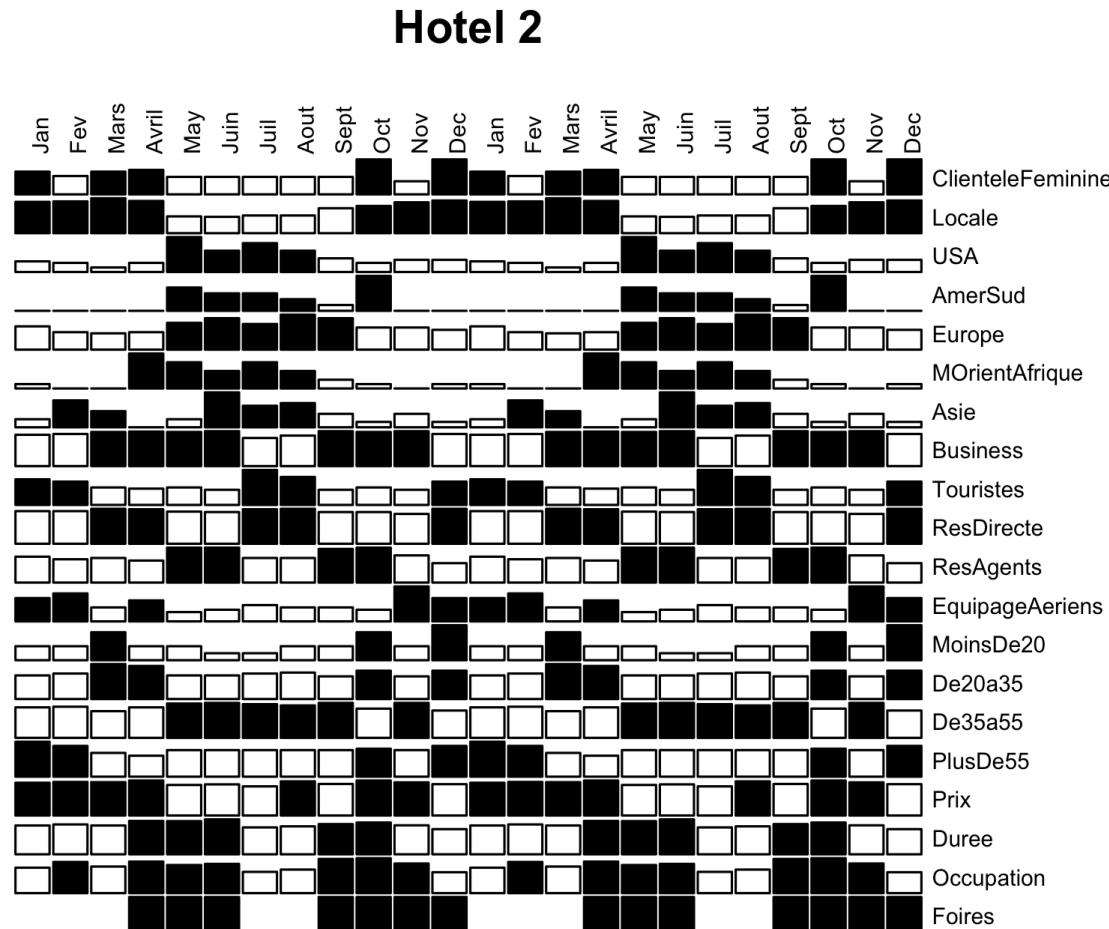
Layouts for multi-dimensional data

- Matrix alignment



Layouts for multi-dimensional data

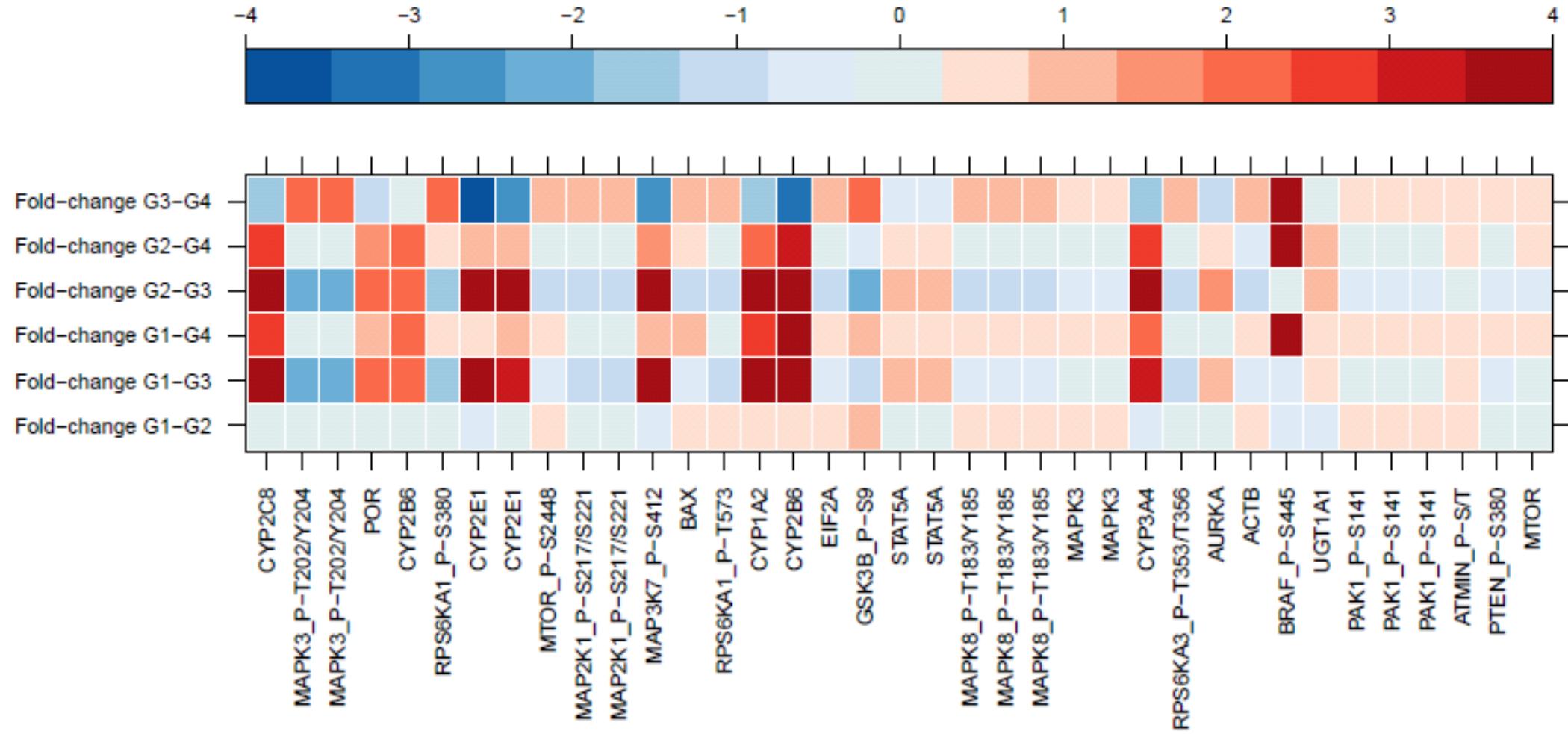
- Matrix alignment. Heat map



Layouts for multi-dimensional data

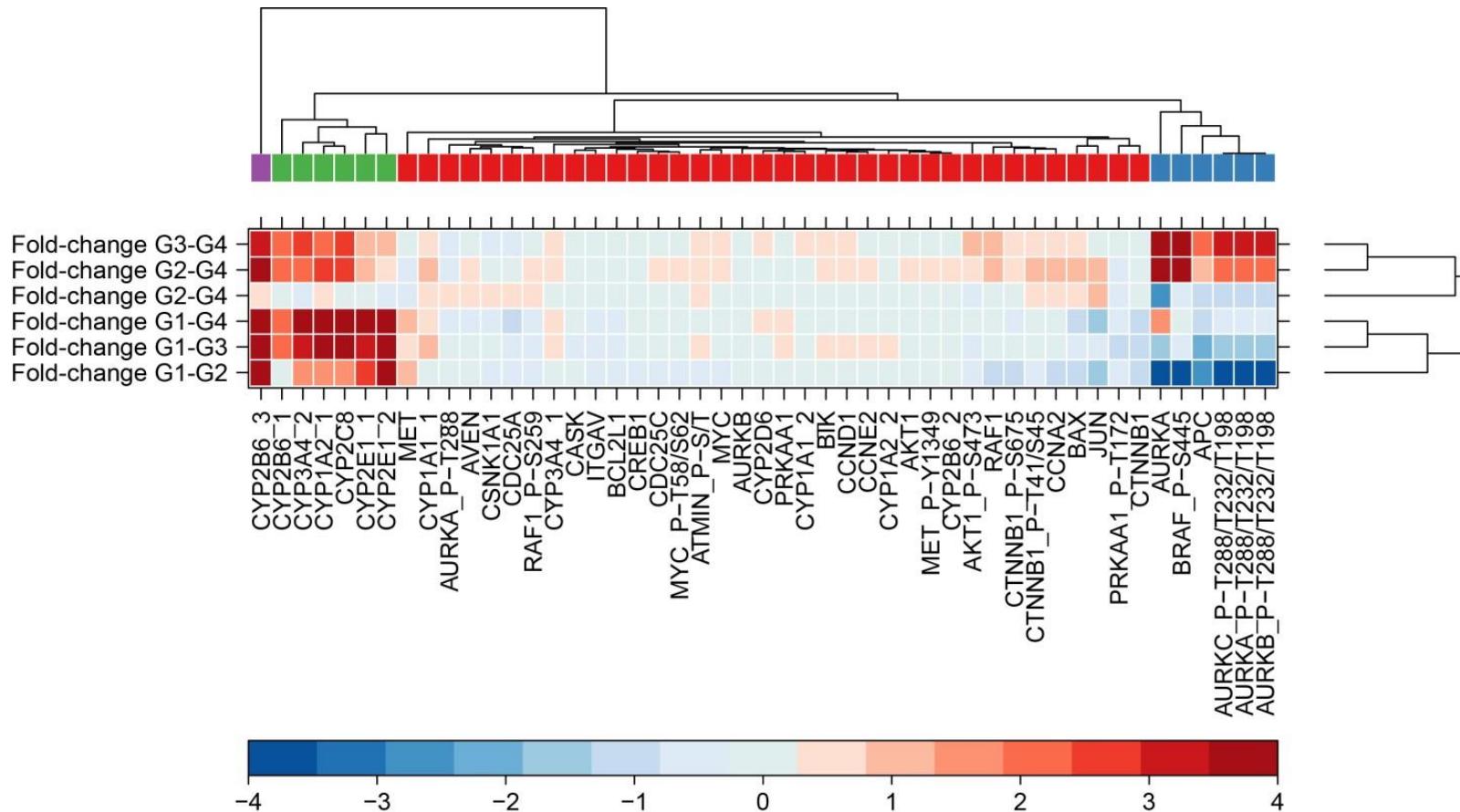
- Matrix alignment. Heat map

<http://www.ra.cs.uni-tuebingen.de/software/RPPApipe/doc/documentation.htm>



Layouts for multi-dimensional data

- Clustered heatmap

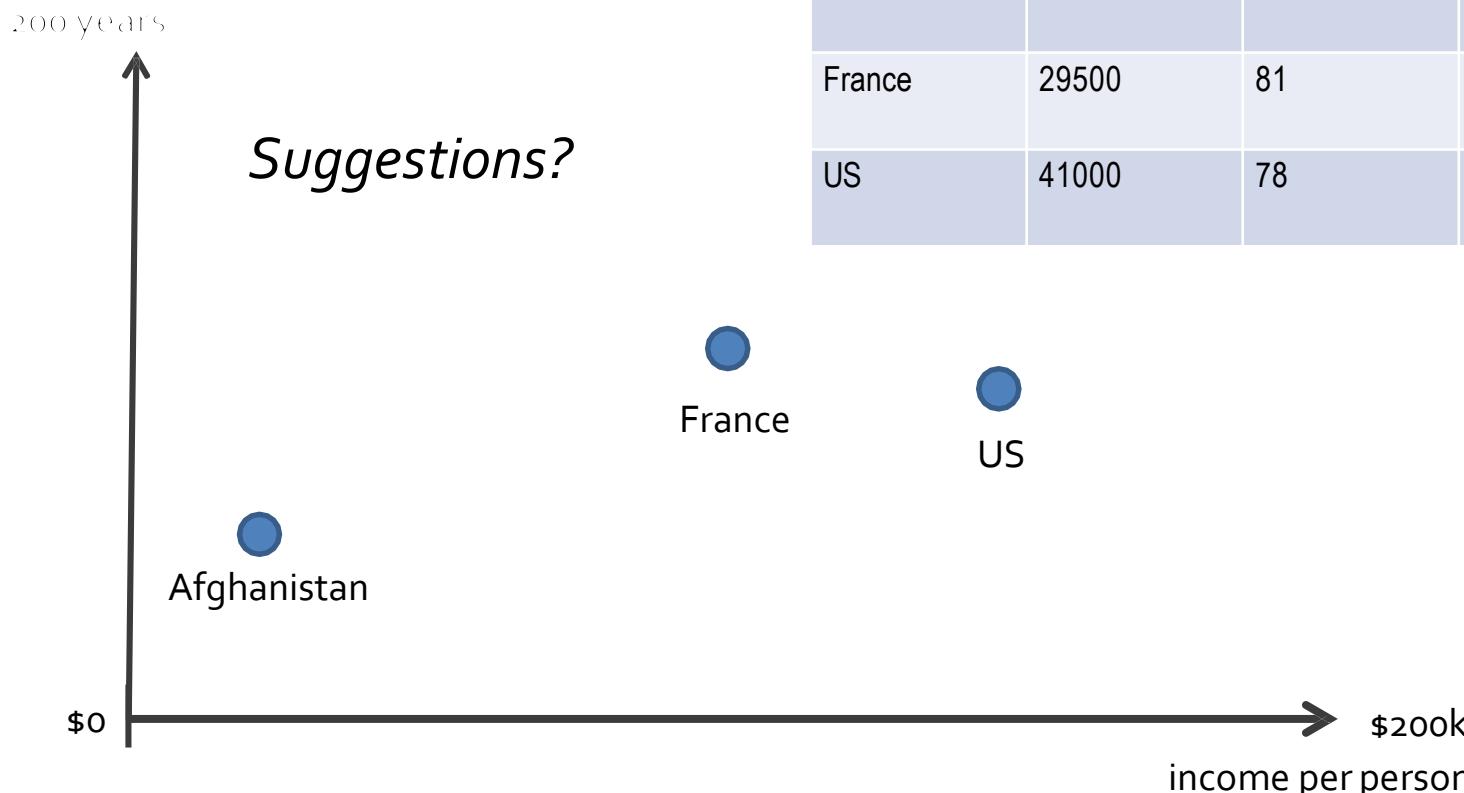


Layouts for multi-dimensional data

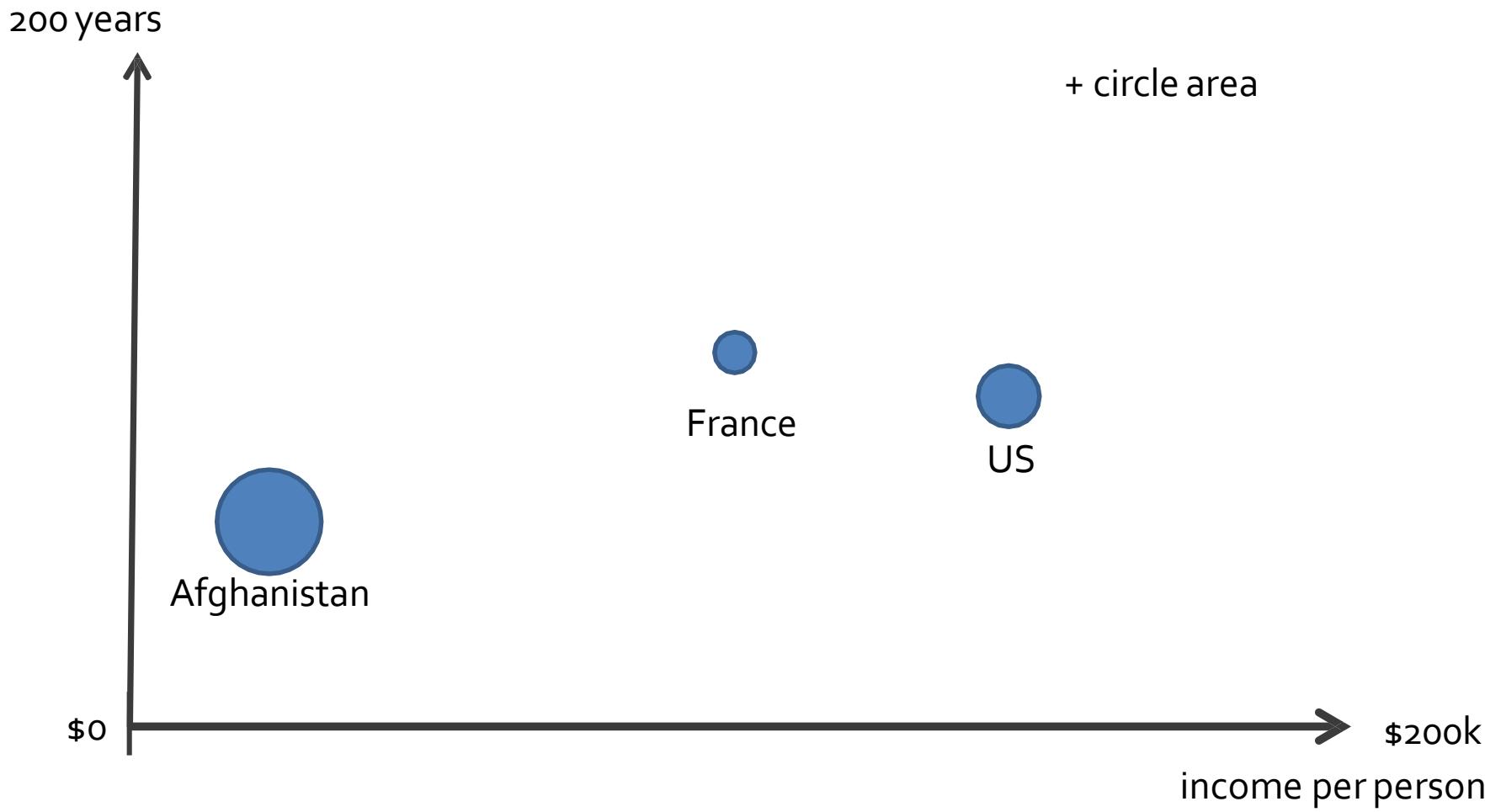
- Increasing attributes in the original example
 - 4 attributes

Country	Income per person	Life expectancy	Children per woman
Afghanistan	850	57	7.1
France	29500	81	1.9
US	41000	78	2.1

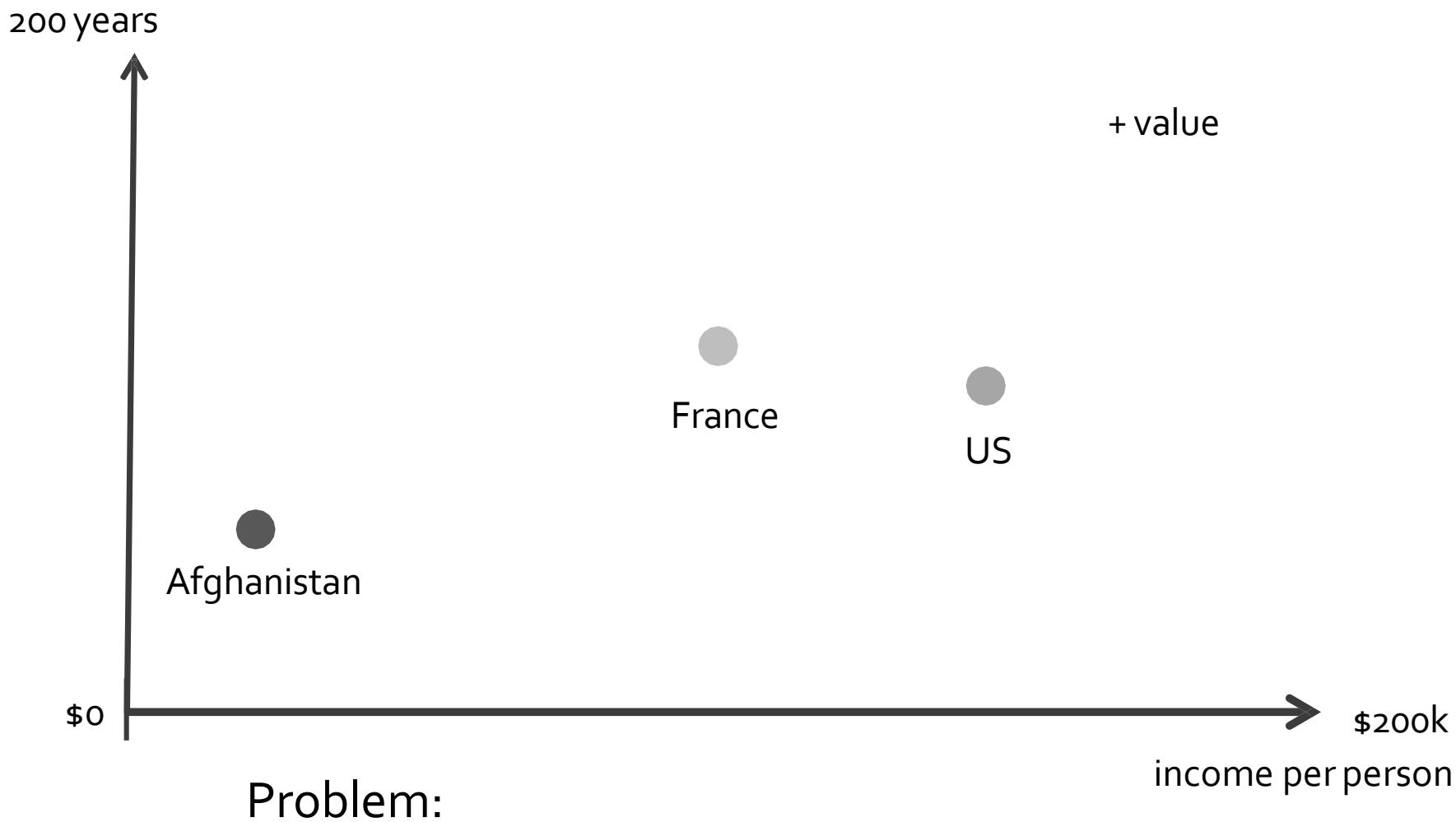
Layouts for multi-dimensional data



Layouts for multi-dimensional data

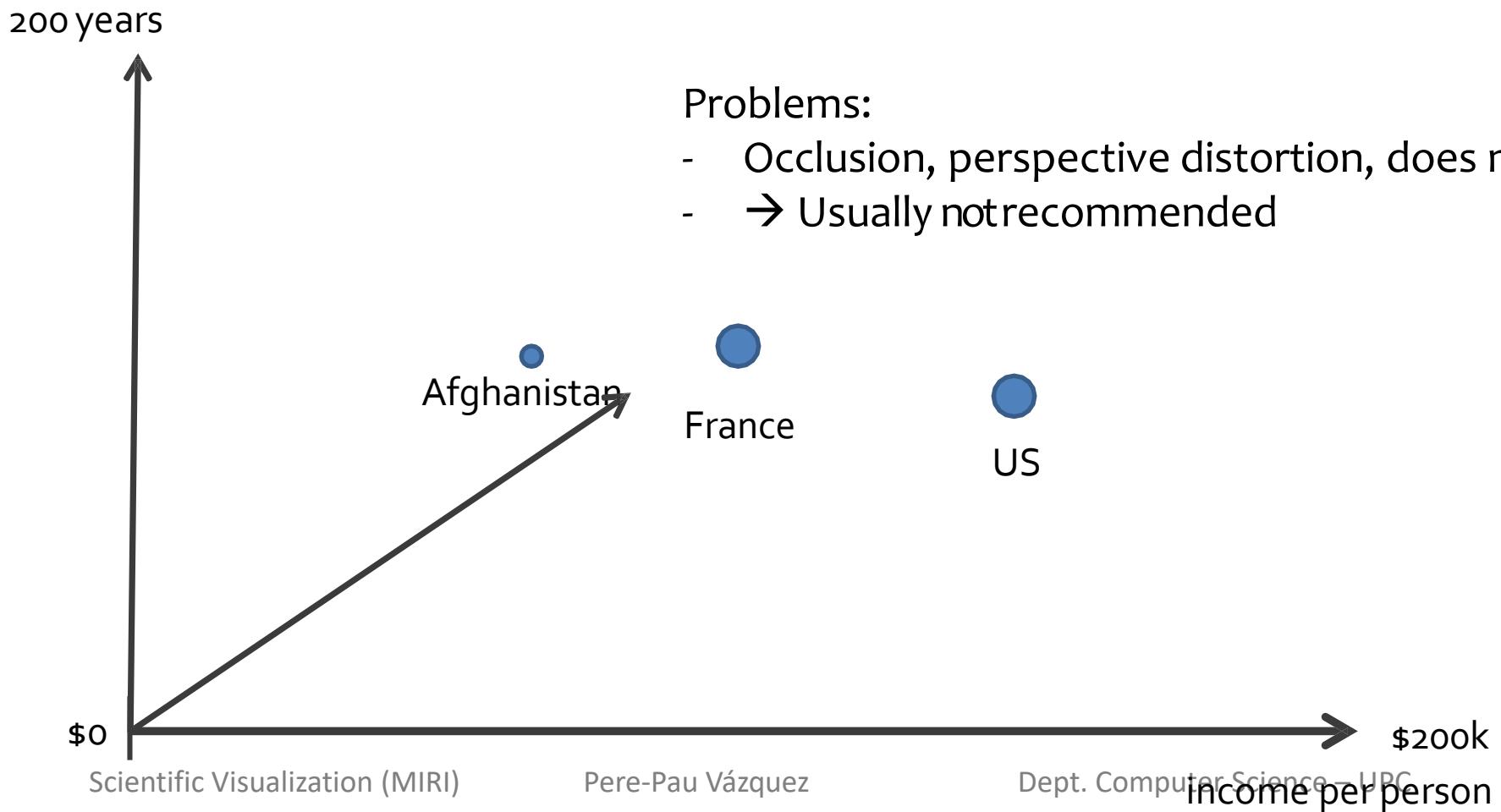


Layouts for multi-dimensional data



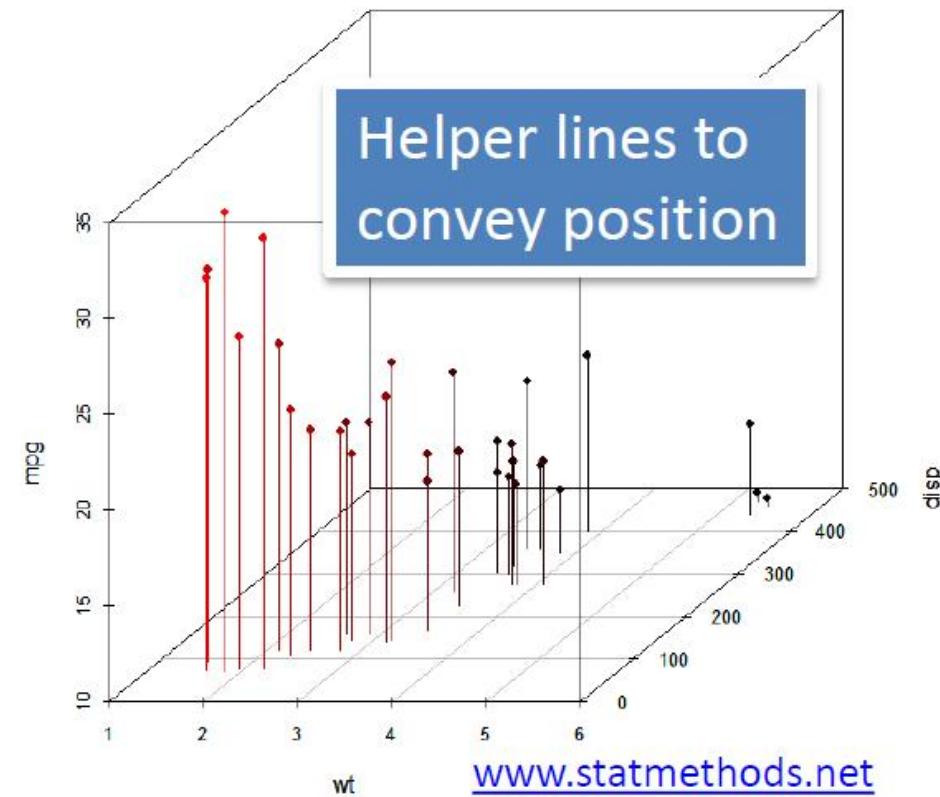
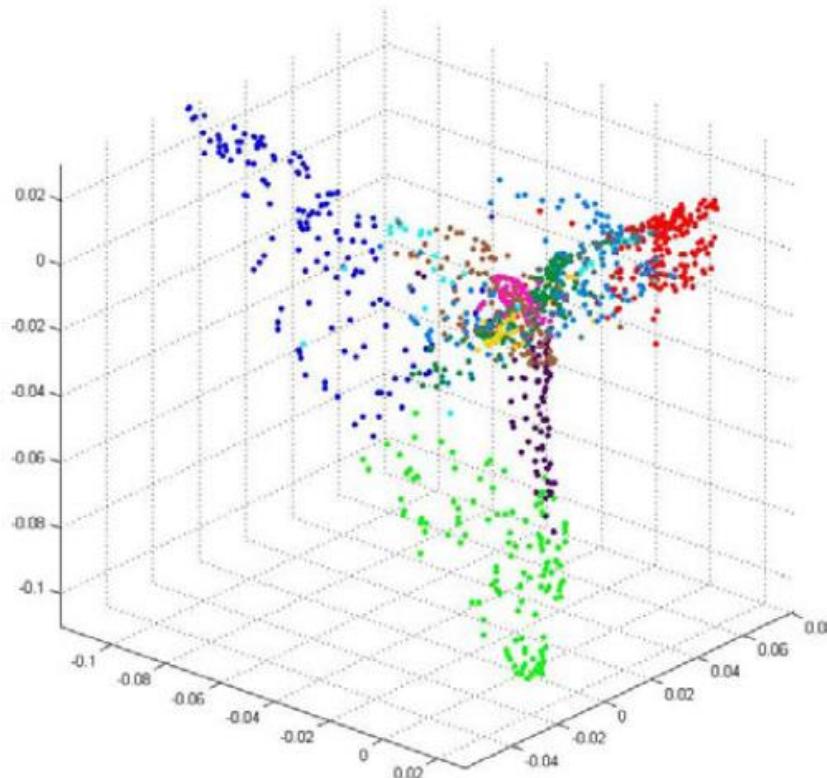
Layouts for multi-dimensional data

- Add an axis



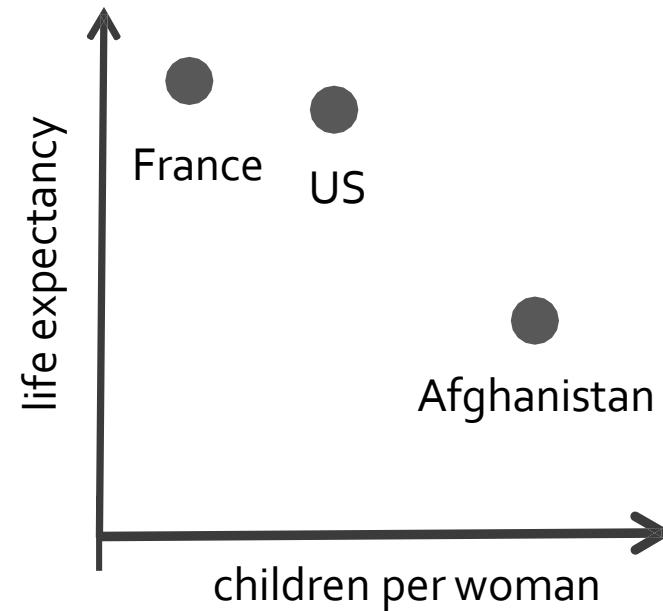
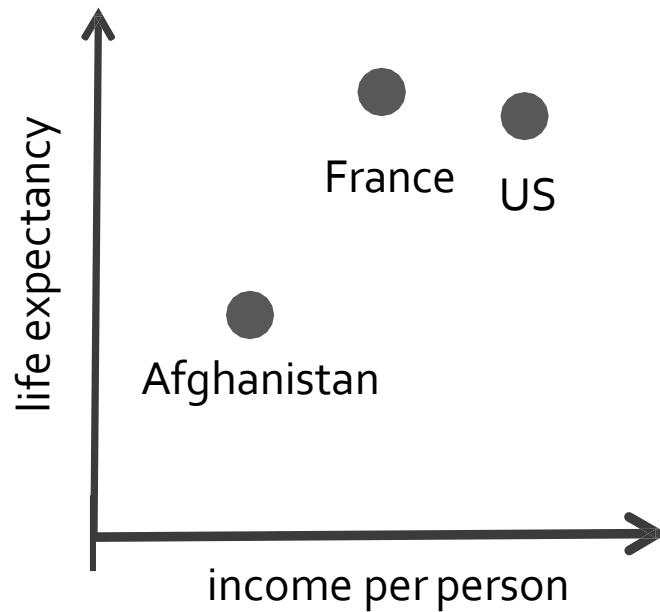
Layouts for multi-dimensional data

- Add an axis: 3D scatterplot
 - Structure, e.g. clusters and point positions still difficult to infer



Layouts for multi-dimensional data

- Add another chart



Layouts for multi-dimensional data

- This idea scales relatively well: scatterplot matrix

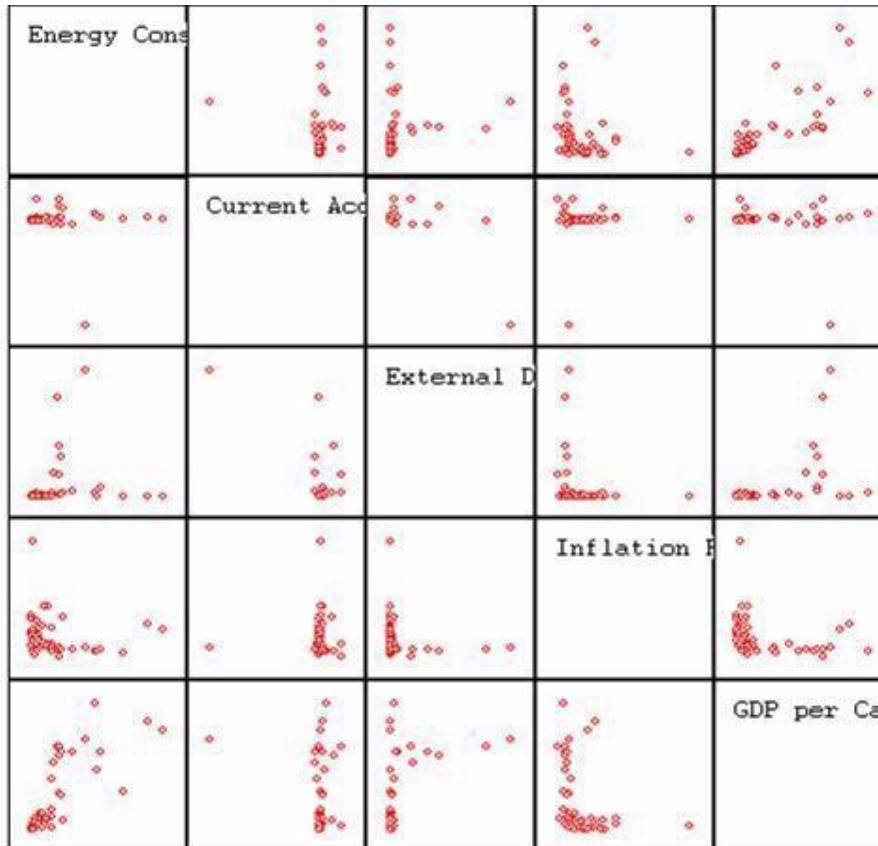


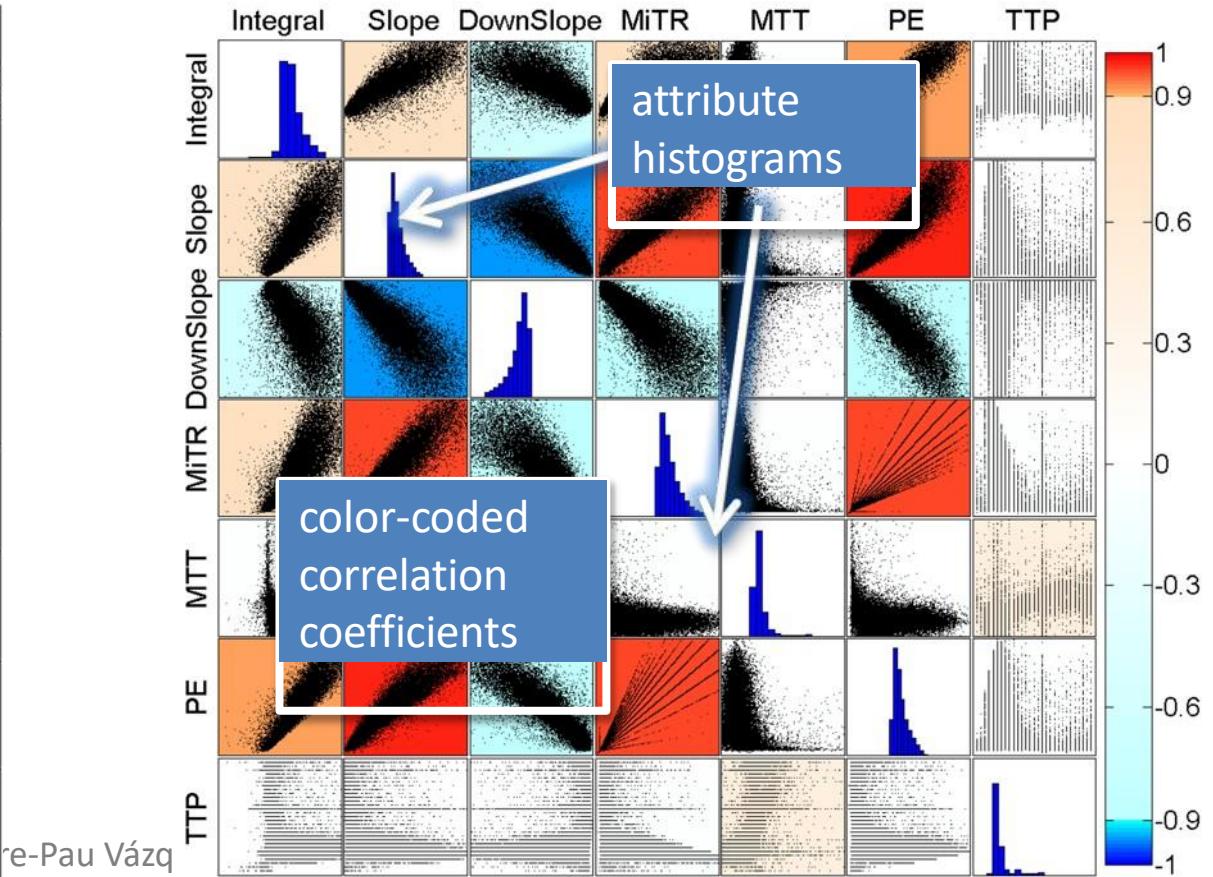
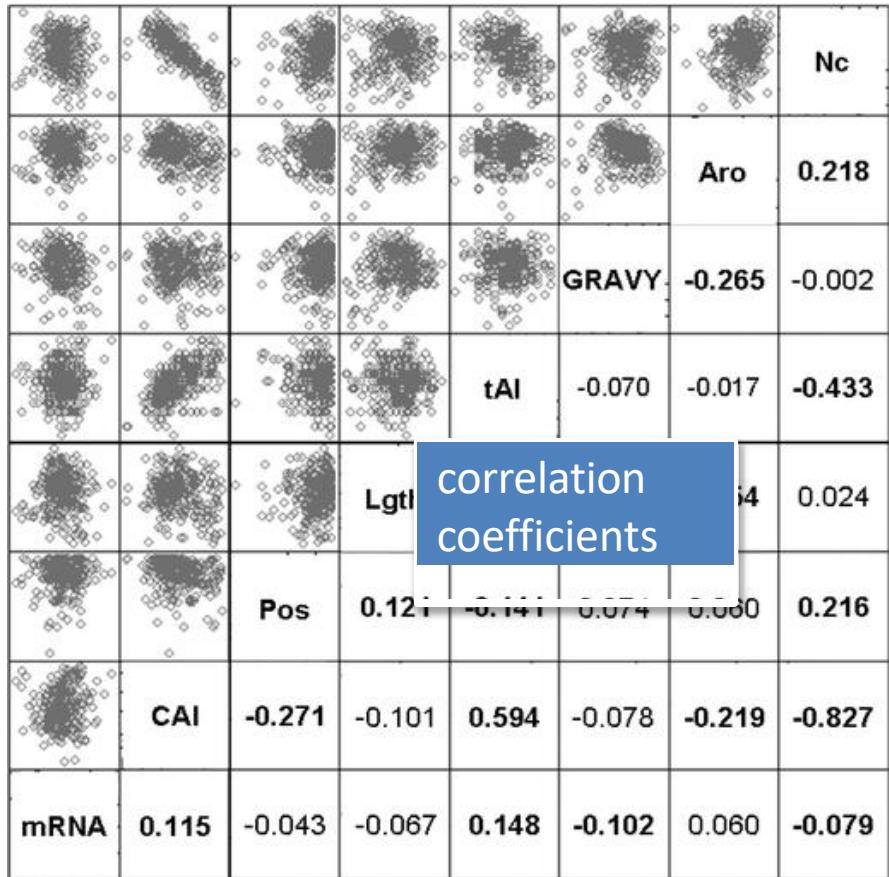
Image Source: Wikipedia

Layouts for multi-dimensional data

- Scatterplot matrix:
 - Arranges scatterplots of all possible attribute pairs
 - Main diagonal and one triangular part of symmetric matrix often show additional information

Layouts for multi-dimensional data

- Scatterplot matrix



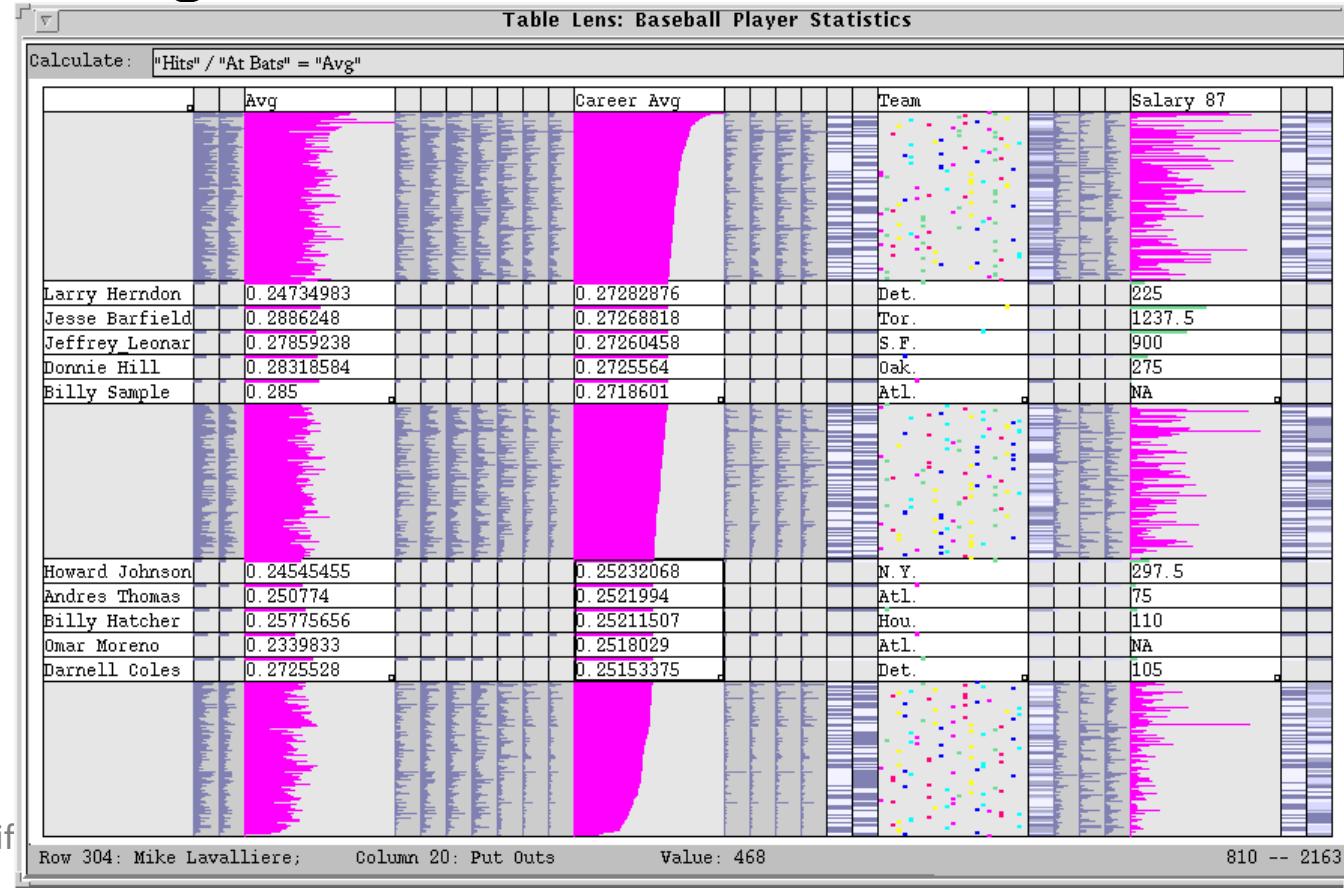
Layouts for multi-dimensional data

- Display the table itself (can be achieved in Excel)
 - Color scales, data bars, symbols

Country Name	GDP per Capita (US\$)	Life Expectancy (Years)	Fertility Rate
Afghanistan	561.1976175	59.6	5.659
Germany	40144.50942	80	1.39
Japan	43117.76827	82.8	1.39
Sierra Leone	447.7533238	44.8	4.943
United States	48357.68451	78.5	1.931
Albania	3764.326348	77	1.741
Algeria	4349.569325	70.6	2.817
American Samoa			
Andorra			1.22
Angola	4218.649126	50.7	6.218
Antigua and Barbuda	13315.24327	75.3	2.13
Argentina	9132.957259	75.7	2.215
Armenia	3124.788199	74.2	1.738
Aruba	24289.14152	75	1.701
Australia	51824.79842	81.7	1.927
Austria	44723.20394	80.4	1.44
Azerbaijan	5843.169753	70.5	1.92
Bahamas, The	21881.08394	74.6	1.901

Layouts for multi-dimensional data

- Display the table itself. With focus+context (table lens)
 - Space advantage of 30 to 100 times

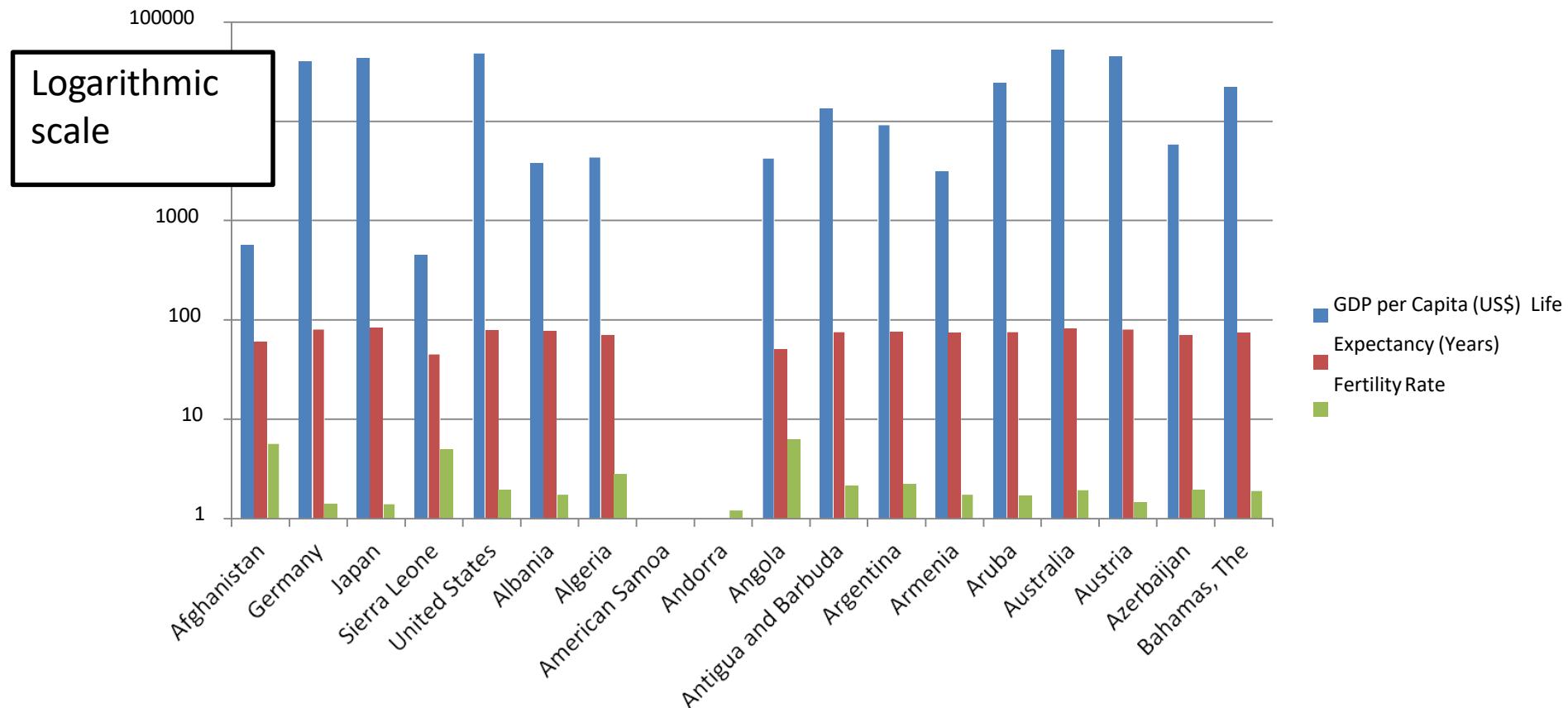


Layouts for multi-dimensional data

- Multiple bar chart: Groups of bars represent multiple attributes
 - Ok for up to 3-5 attributes
 - Though problems for strongly varying scales

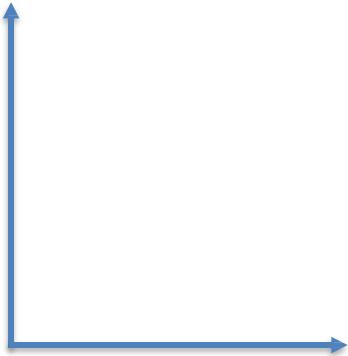
Layouts for multi-dimensional data

- Multiple bar chart

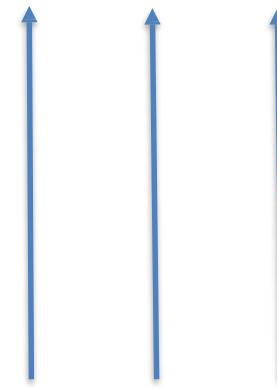


Layouts for multi-dimensional data

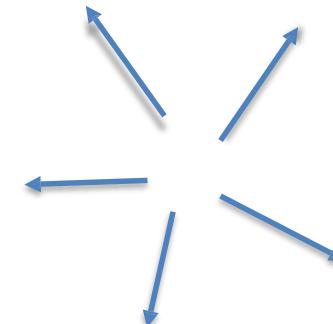
- Changing axis orientation



rectilinear



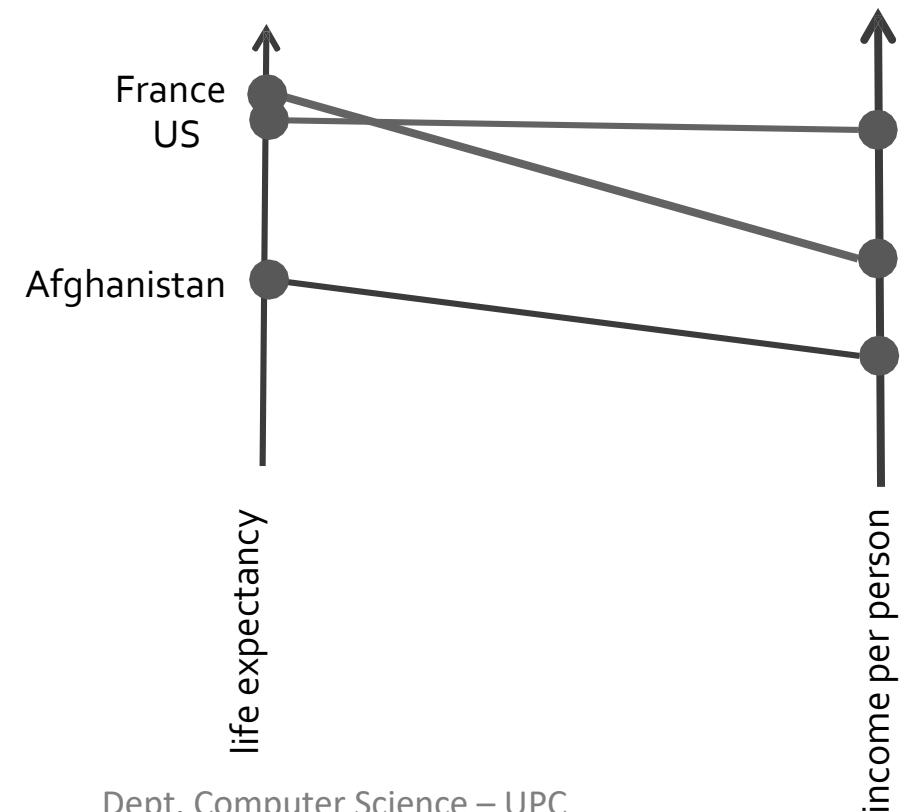
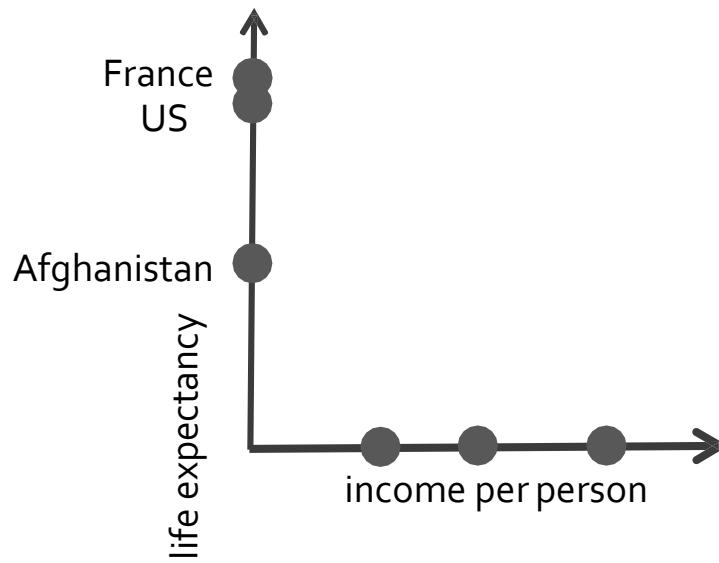
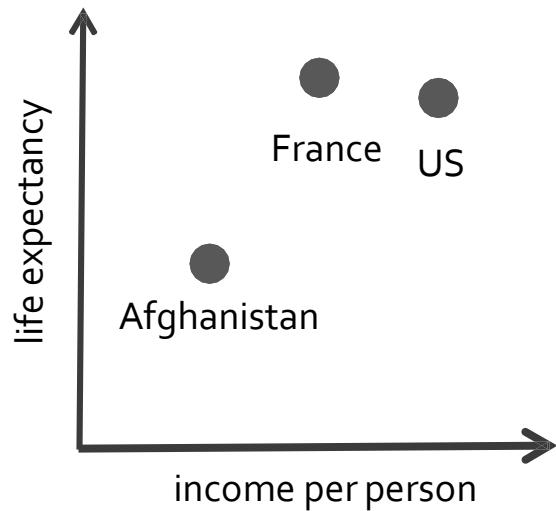
parallel



radial

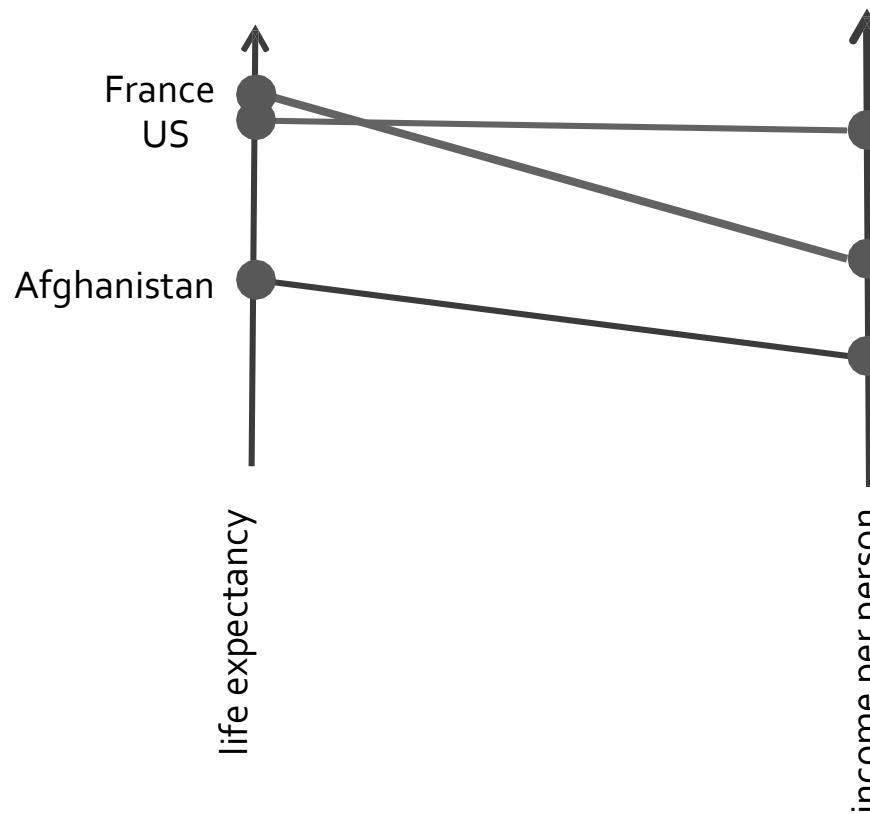
Layouts for multi-dimensional data

- Back to our original example

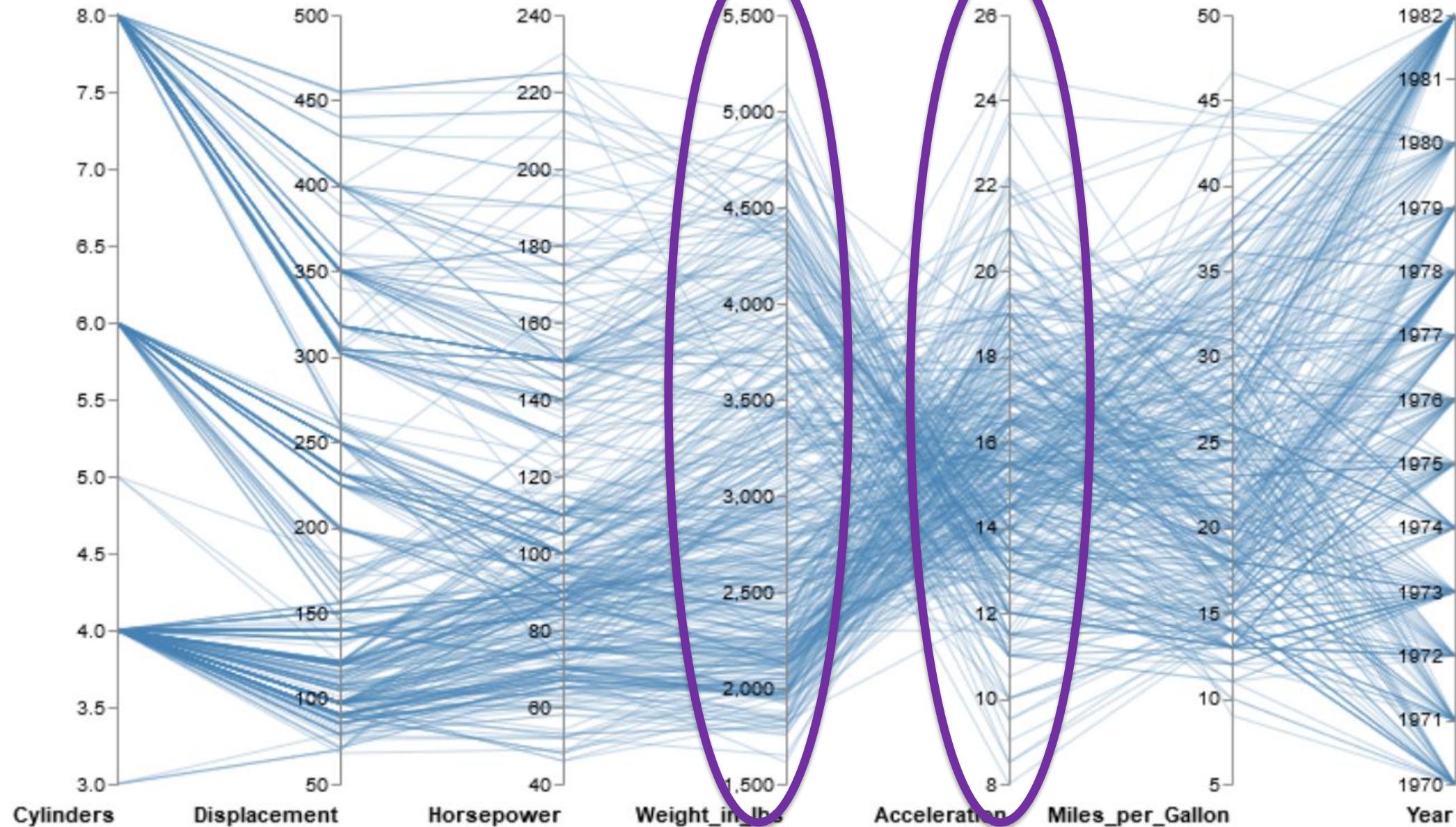


Layouts for multi-dimensional data

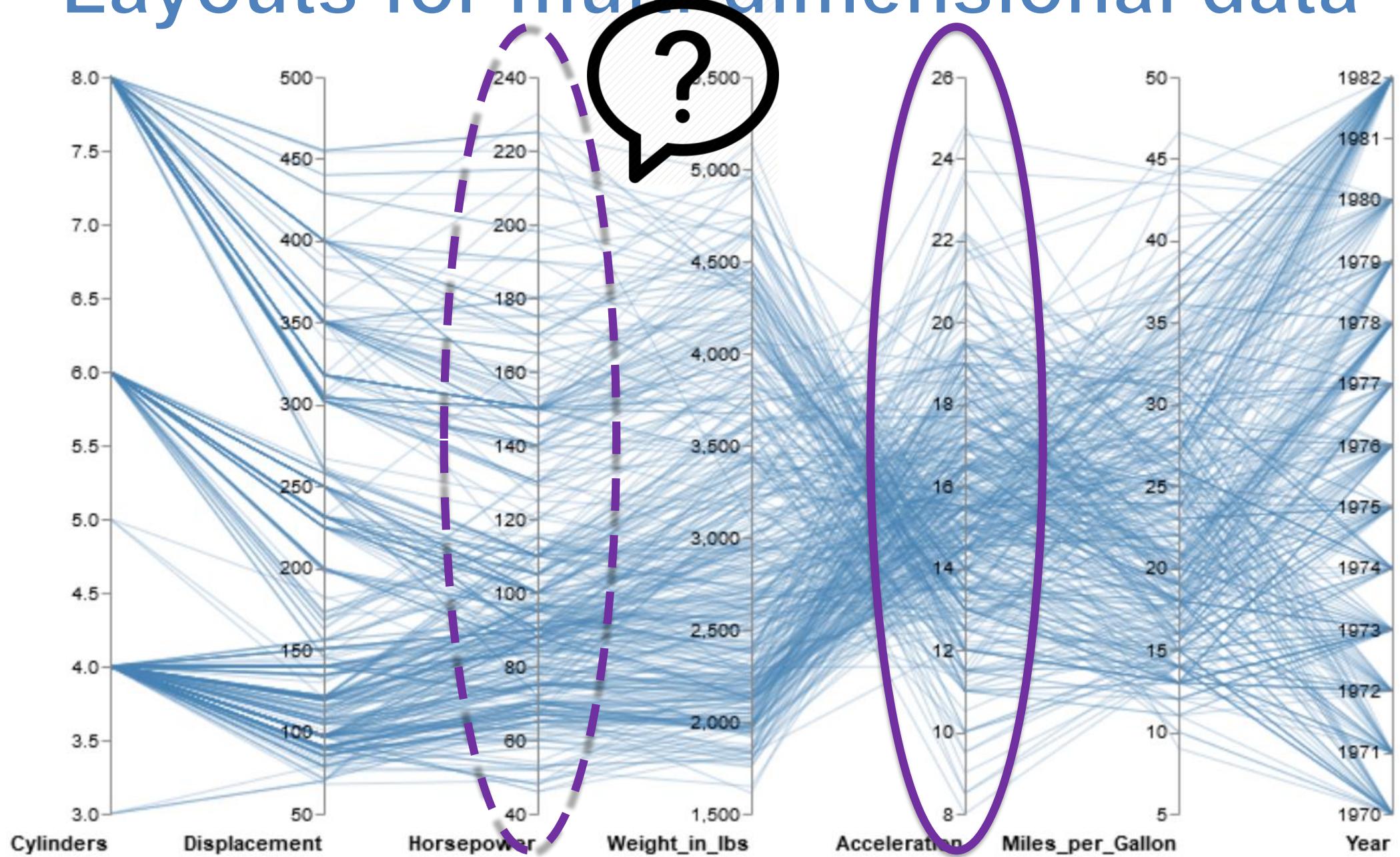
- Parallel coordinates: show correlations between neighboring axes



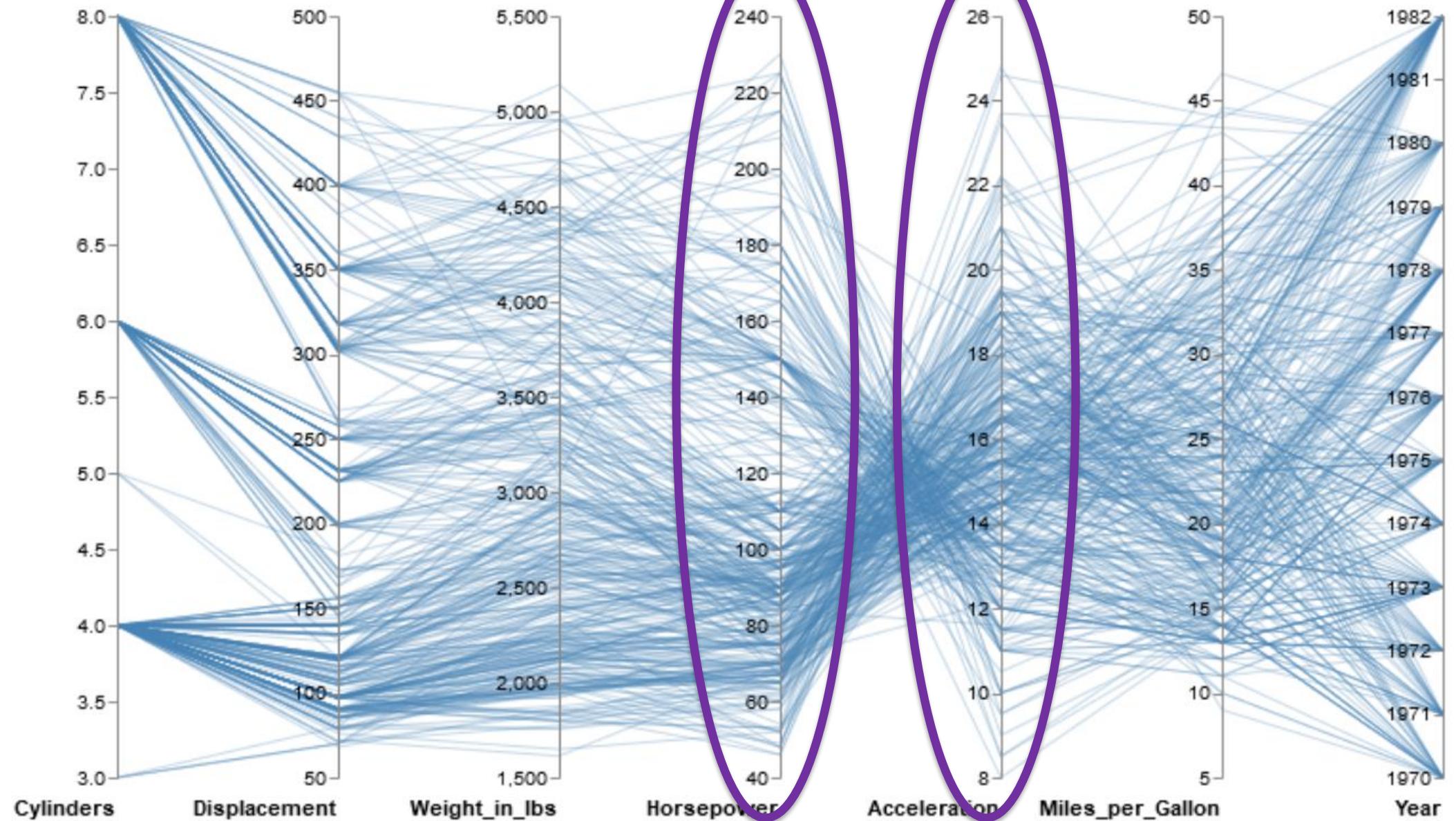
Layouts for multi-dimensional data

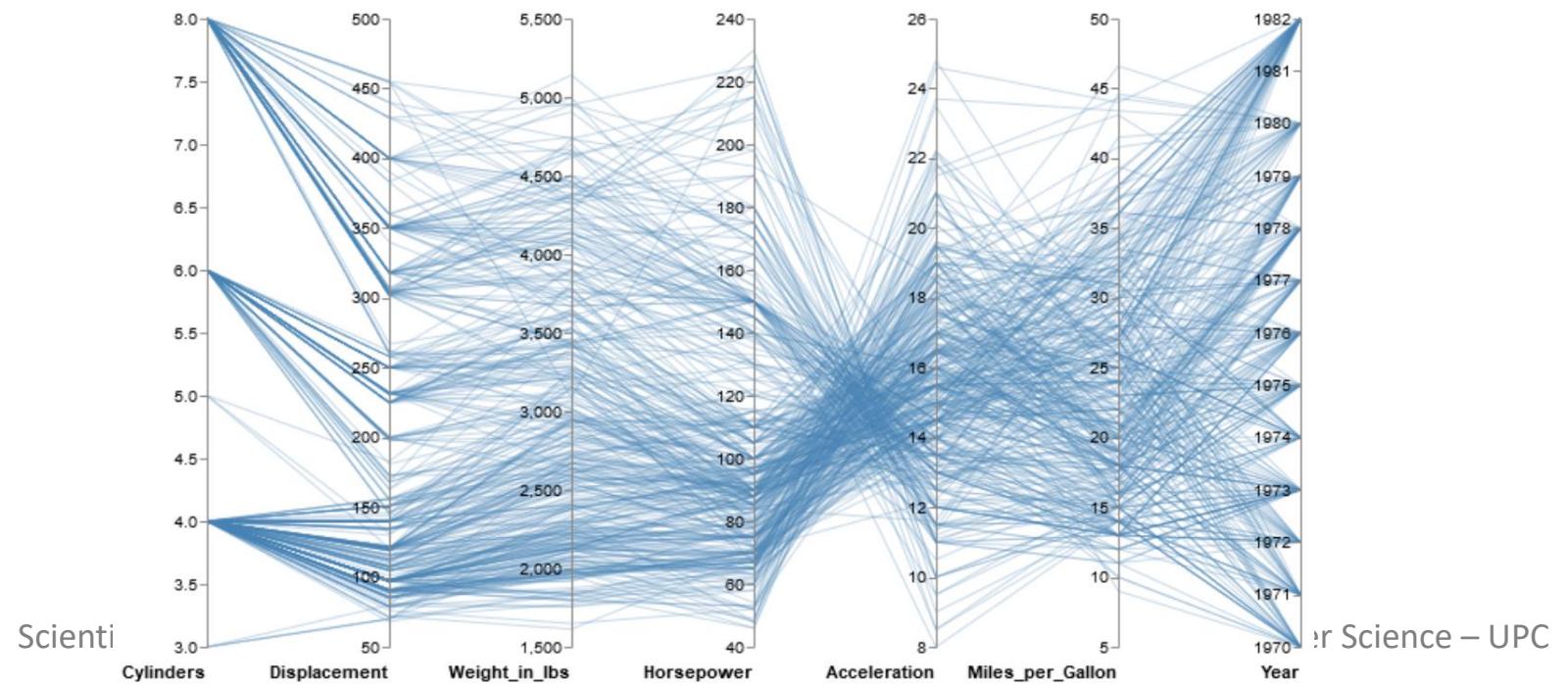
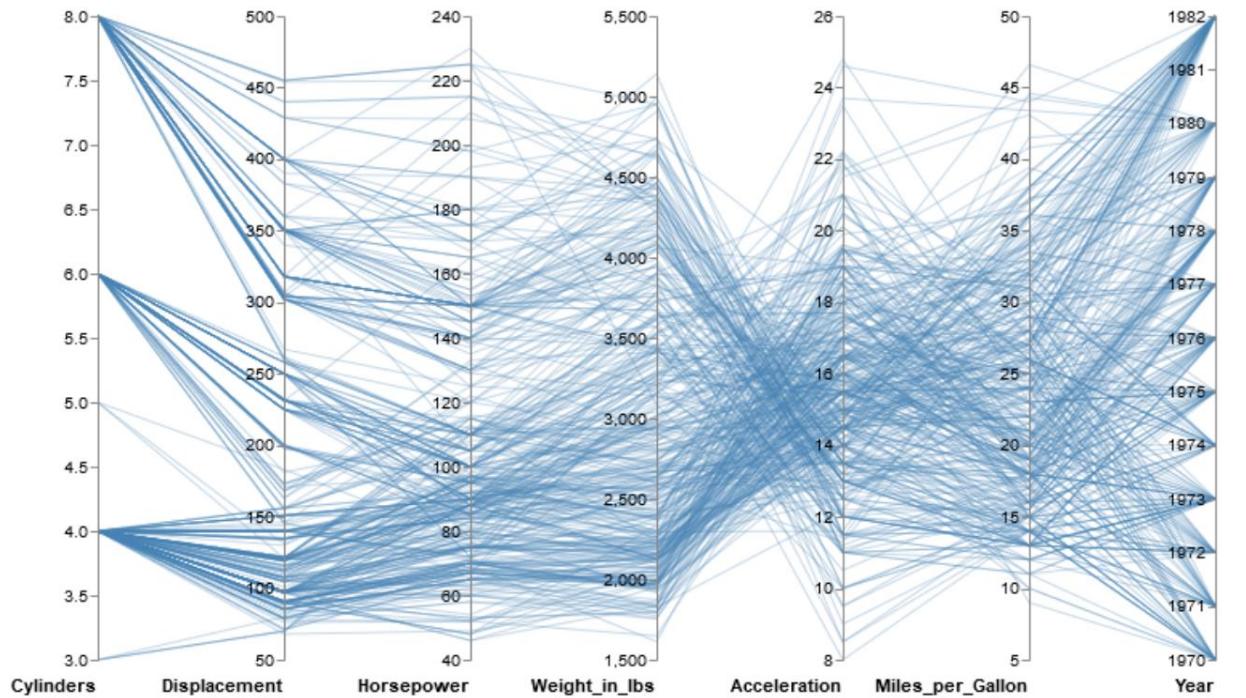


Layouts for multi-dimensional data

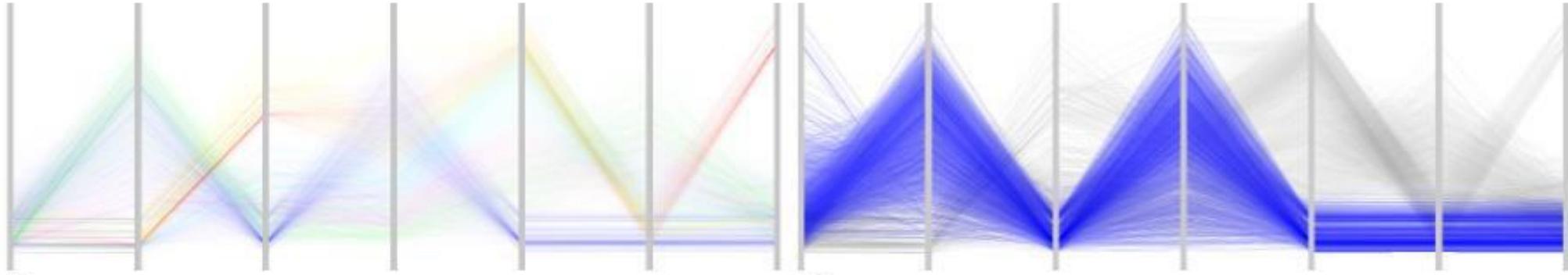


Layouts for multi-dimensional data



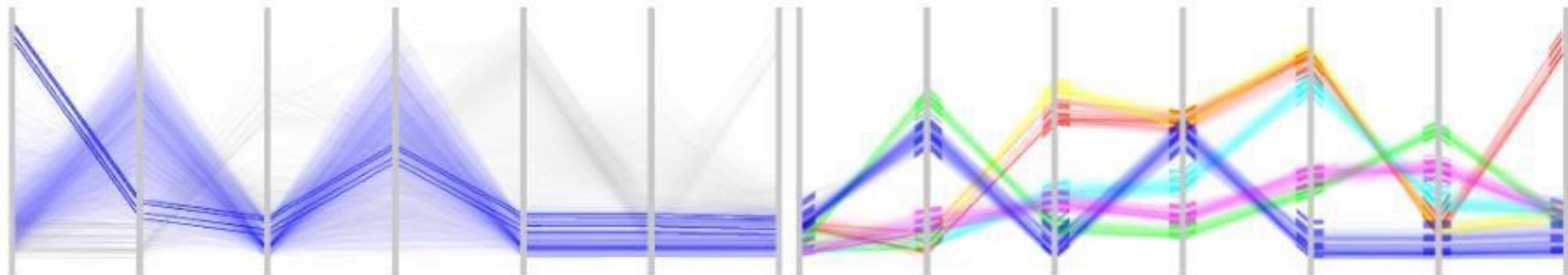


Layouts for multi-dimensional data



(a) A linear transfer function has been applied to the high-precision texture in order to prevent cluttering and to provide overview of the data.

(b) A logarithmic transfer function is applied to a selected cluster. The structure is preserved and emphasis is put on the low density regions.

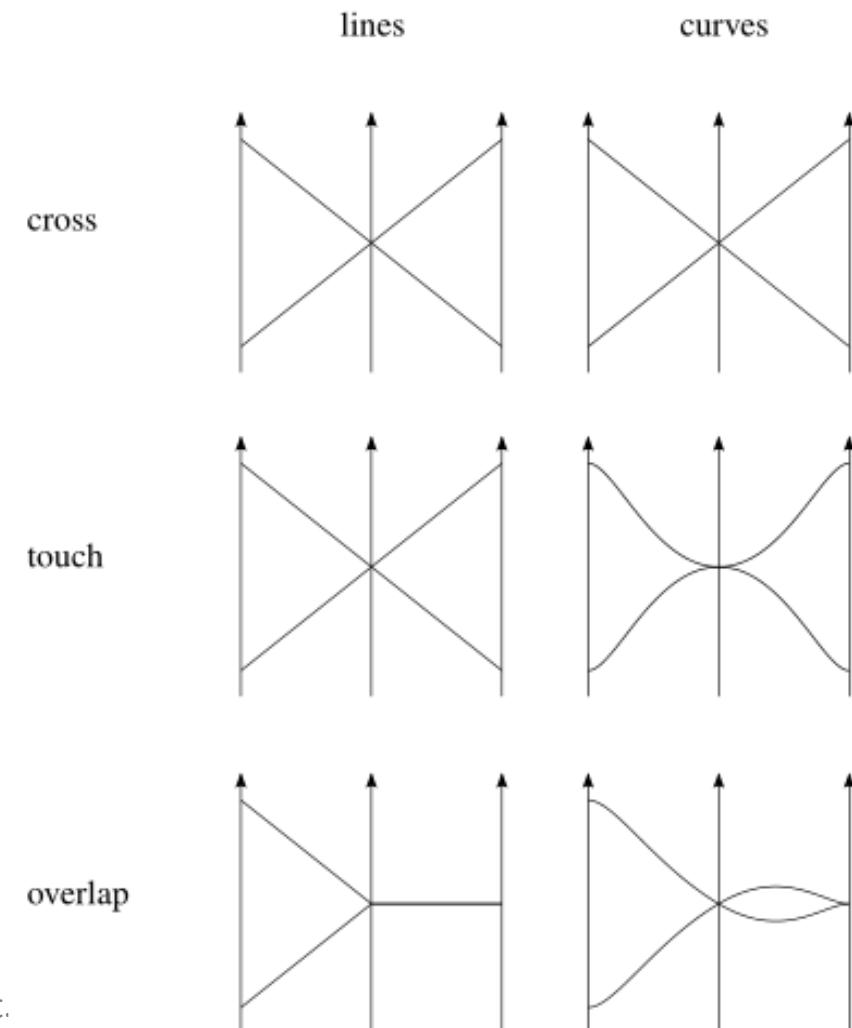


(c) Local cluster outliers are enhanced. A square root transfer function is used and the outliers are visible even through high-density regions.

(d) A complementary view of the clusters with uniform bands. 'Feature animation' presents statistics about the clusters and acts as a guidance.

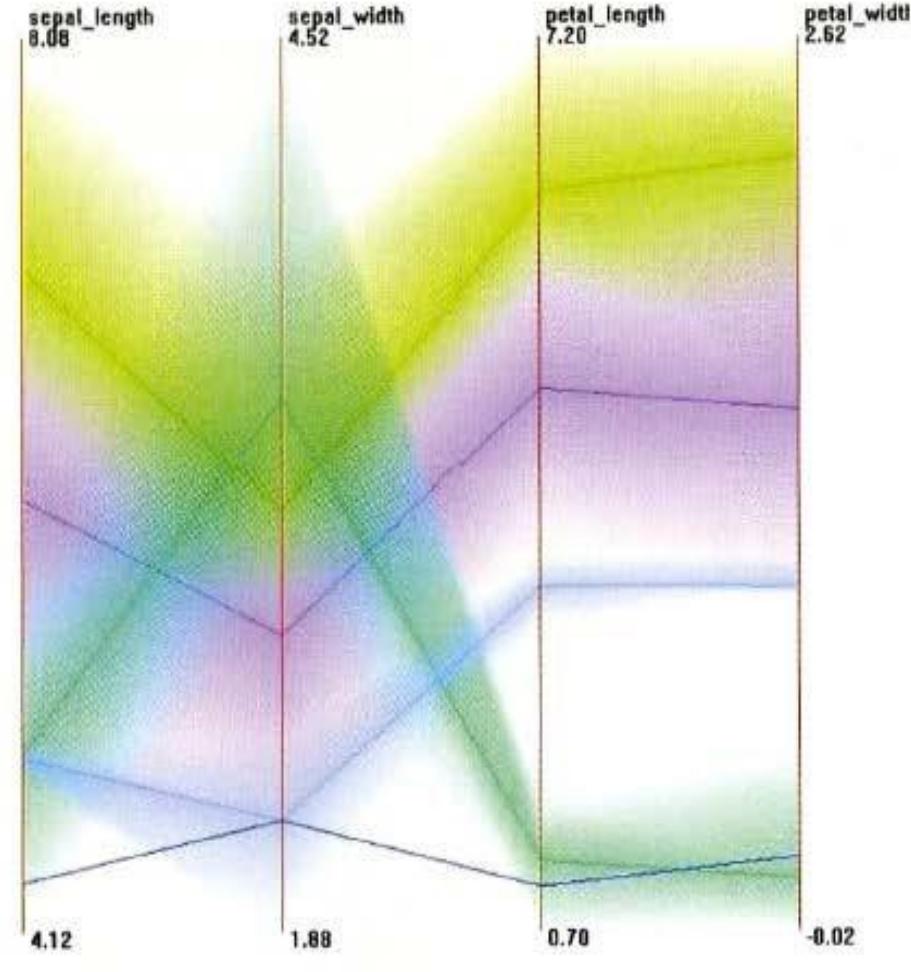
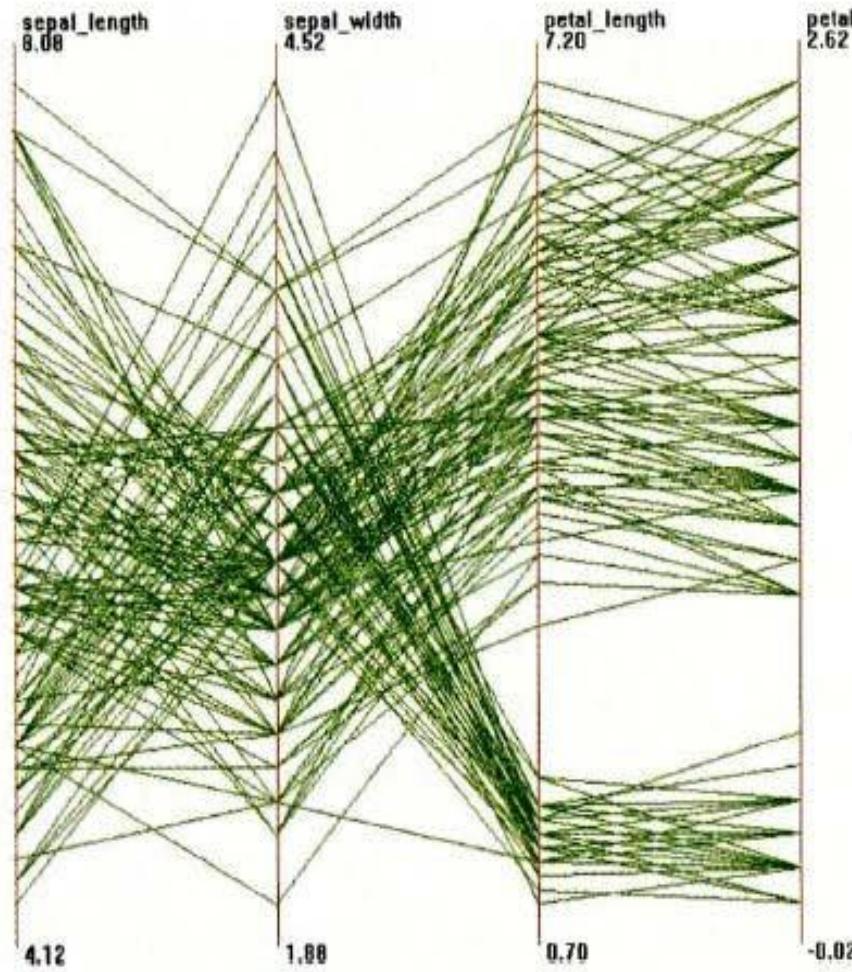
Layouts for multi-dimensional data

- How to draw parallel coordinate lines?
 - Try to avoid ambiguity



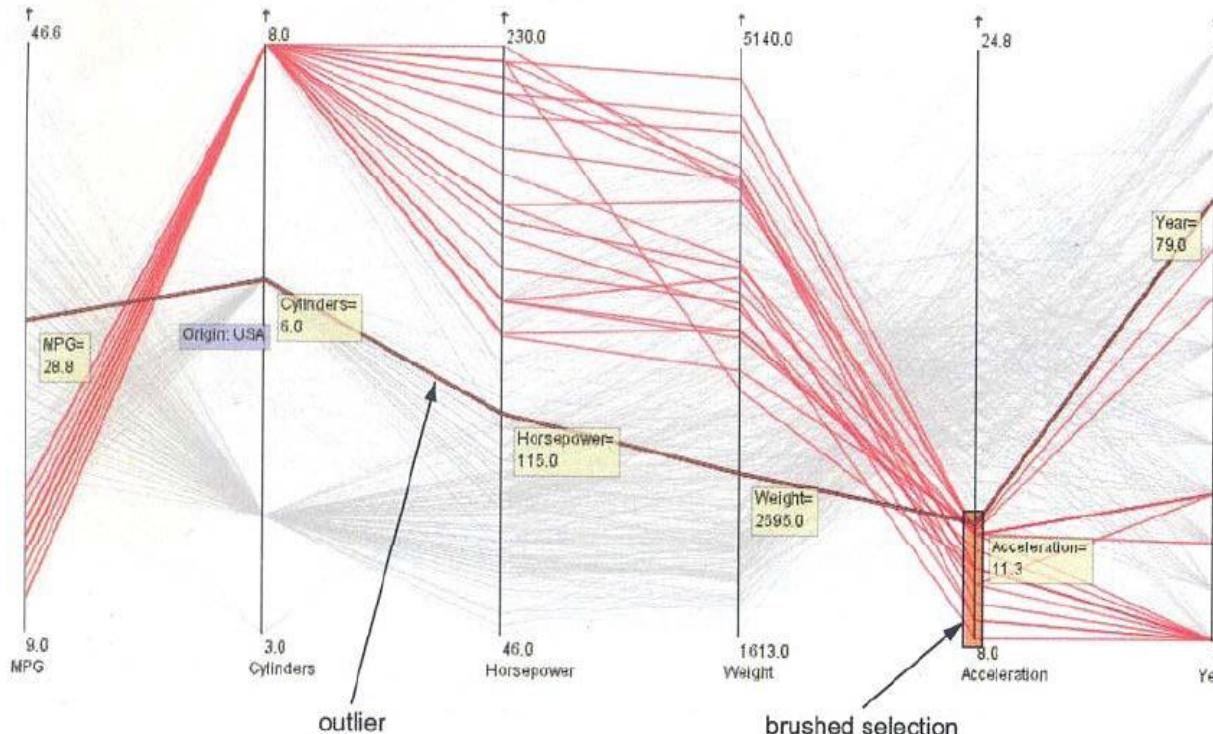
Layouts for multi-dimensional data

- Clustering in parallel coordinates



Layouts for multi-dimensional data

- Item selection in Parallel coordinates
 - Need to brush a single or multiple axis



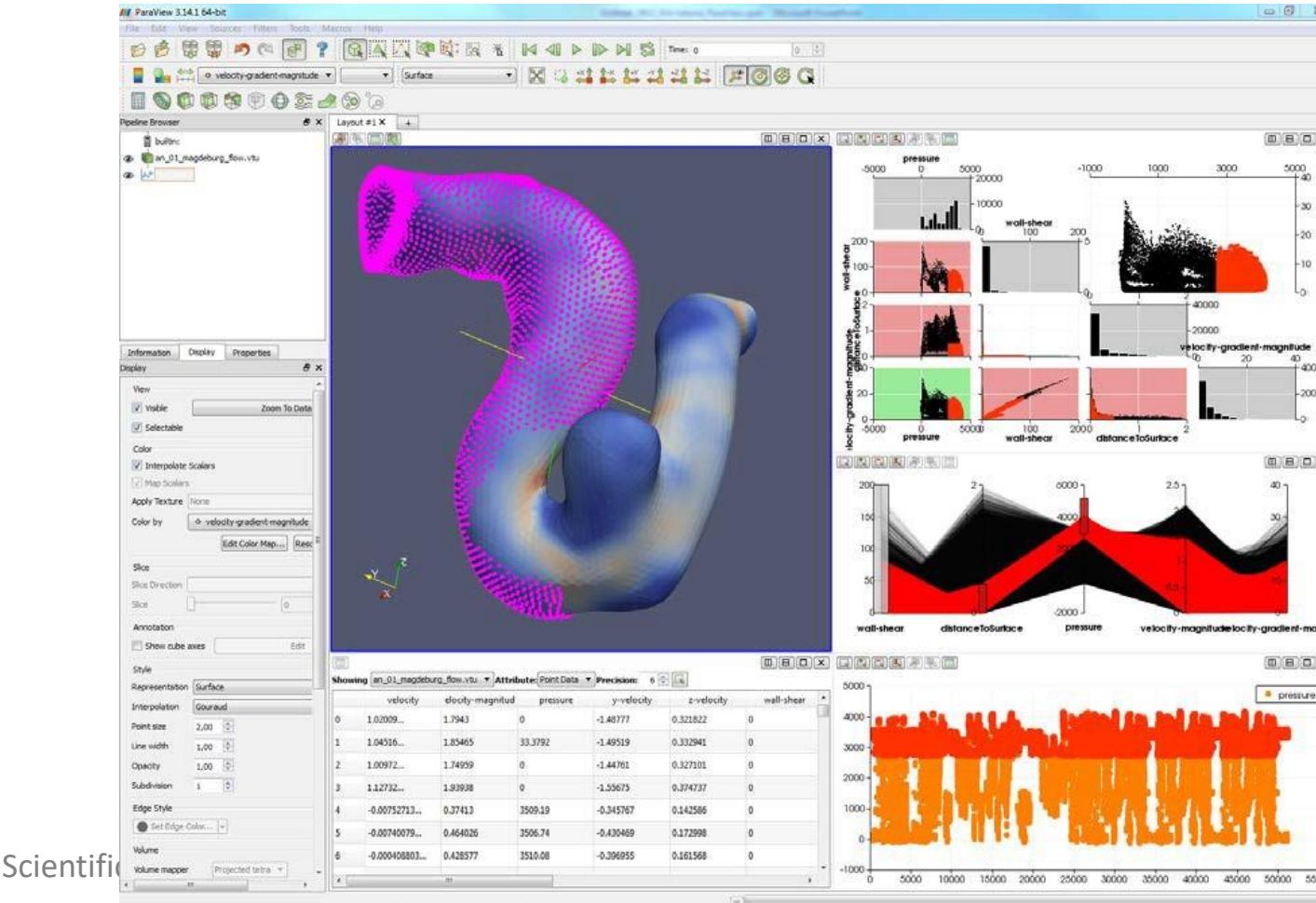
[Telea08]

Layouts for multi-dimensional data

- Multiple coordinated views
 - One dataset but multiple views on the data
 - Views are linked and equipped with brushing facilities

Layouts for multi-dimensional data

- Multiple coordinated views



Implemented in [ParaView](#)

Layouts for multi-dimensional data

- So, what type of visualization?
 - Depends on the data
 - Depends on the tasks
 - ...

Chart Suggestions—A Thought-Starter

www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

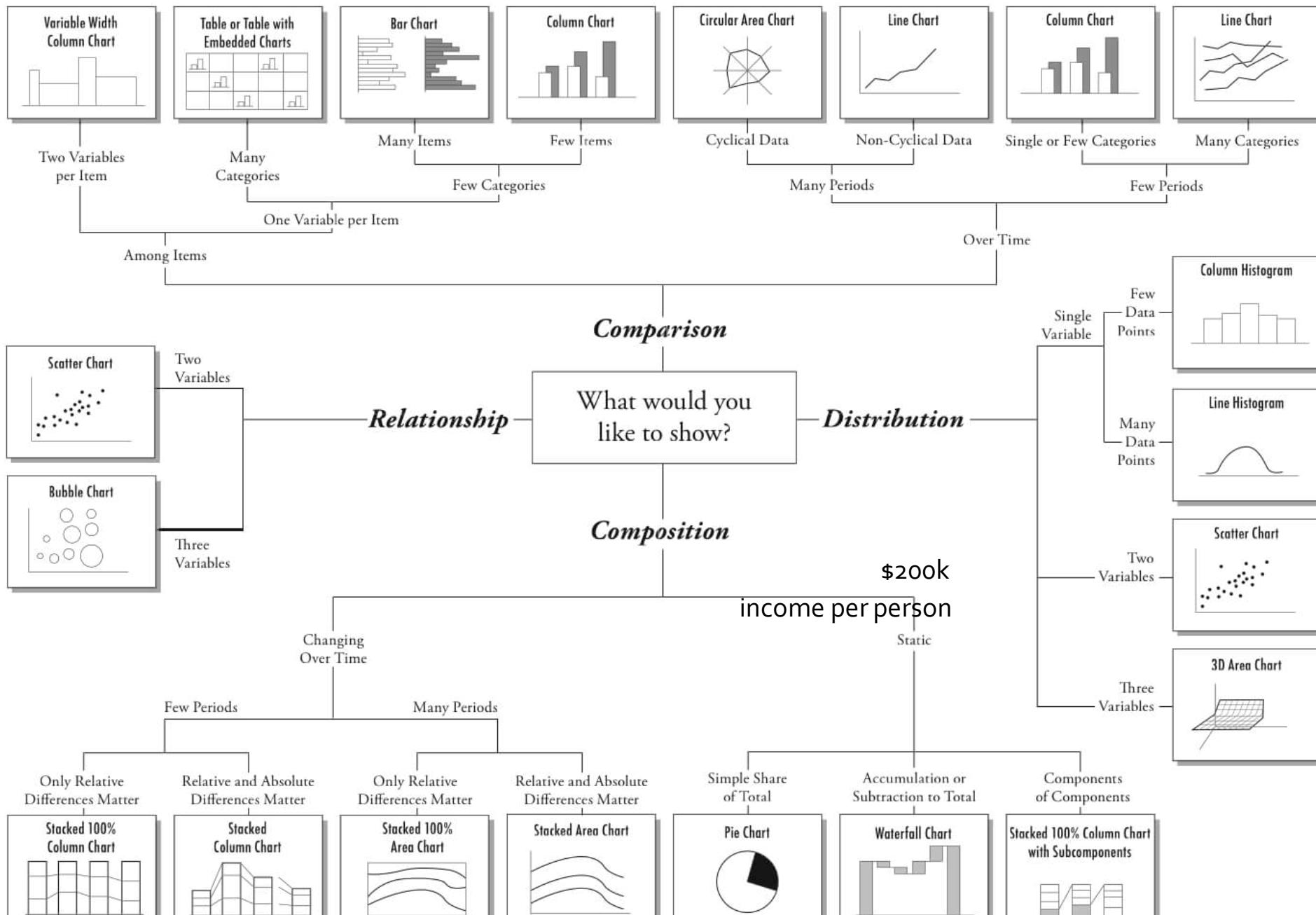


Chart Suggestions—A Thought-Starter

www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

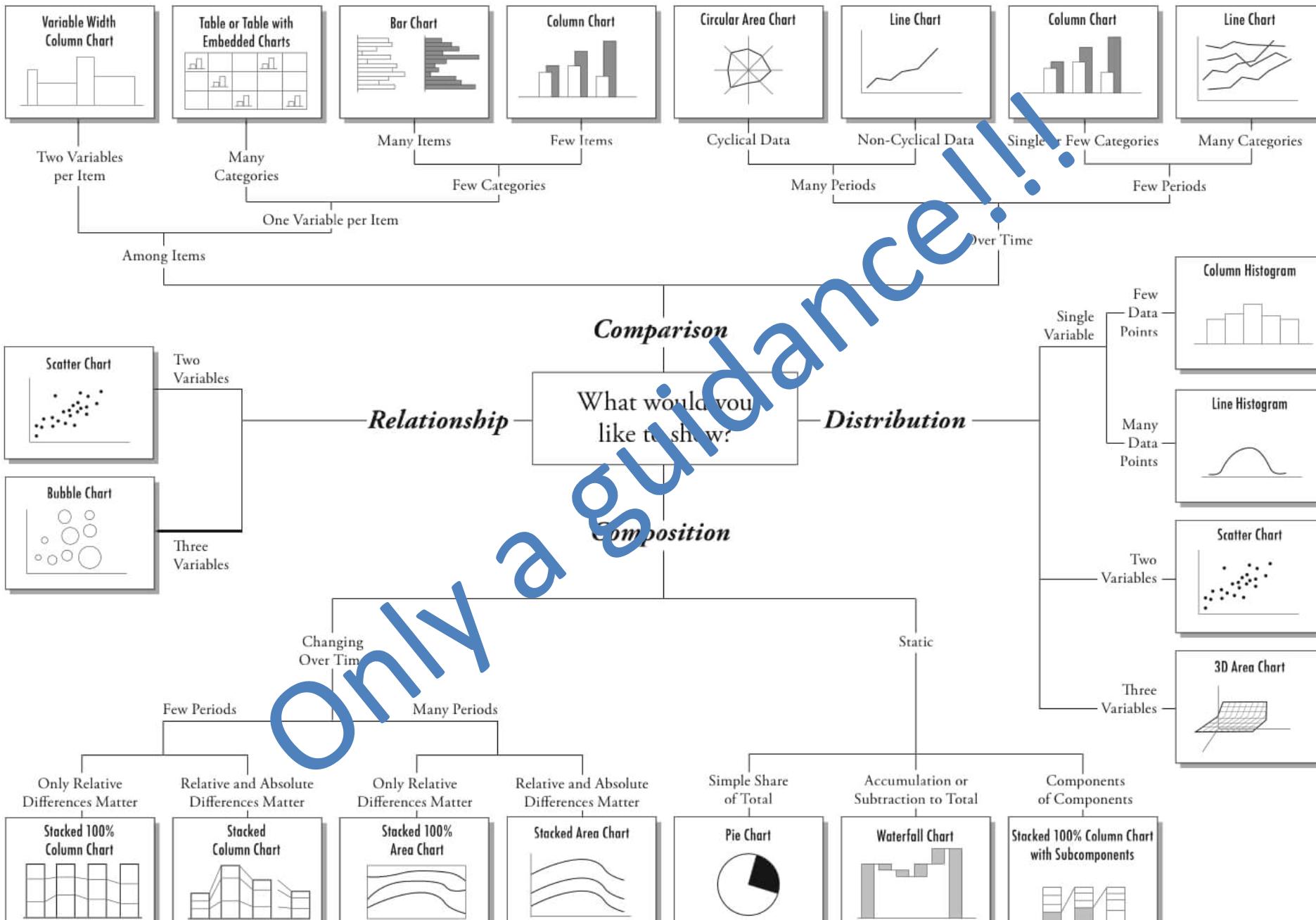
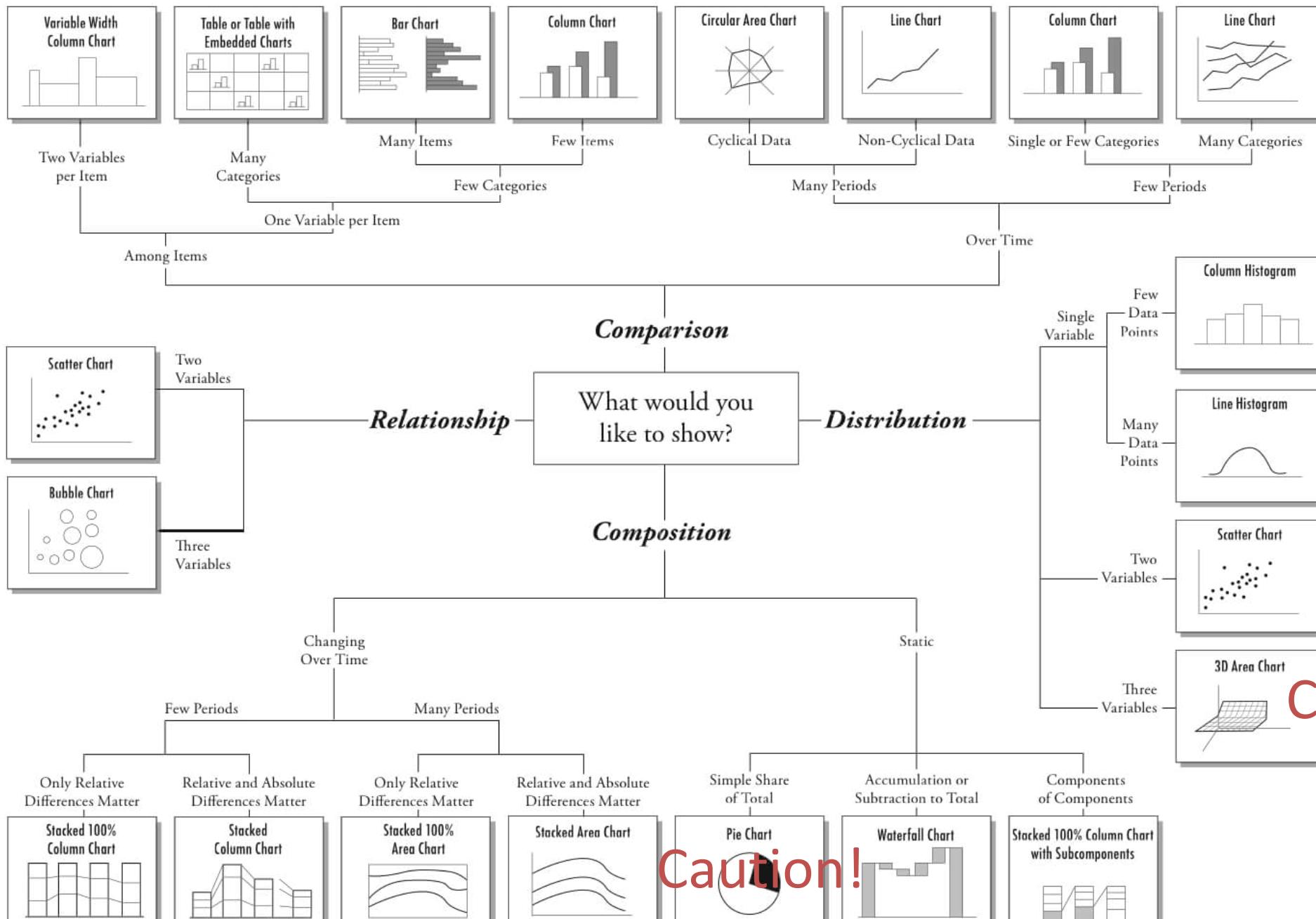


Chart Suggestions—A Thought-Starter

www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com



Visualization. Multi- Dimensional Data Visualization

Pere-Pau Vázquez
Dept. Computer Science – UPC