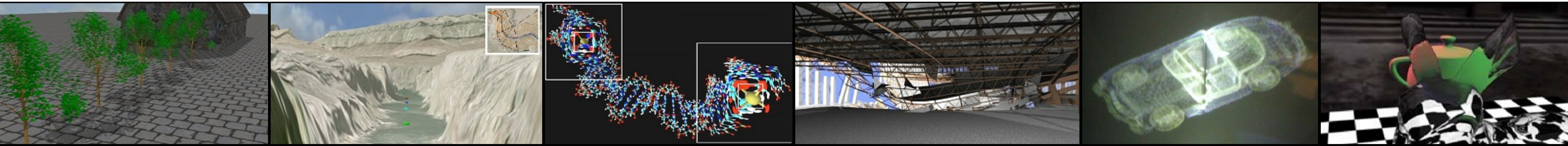
CIS 4930/6930-002
DATA VISUALIZATION

TABULAR DATA

Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah)

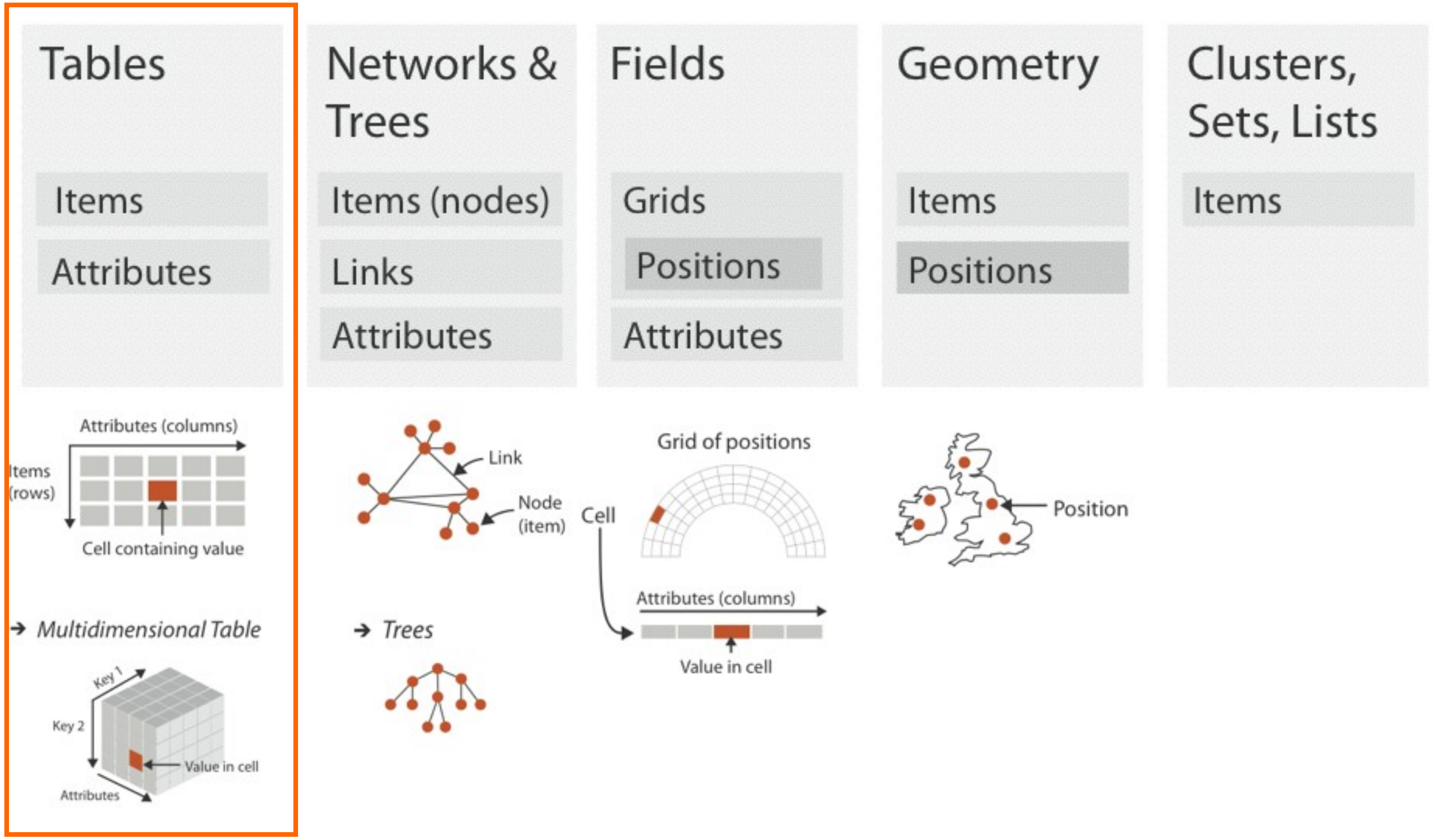


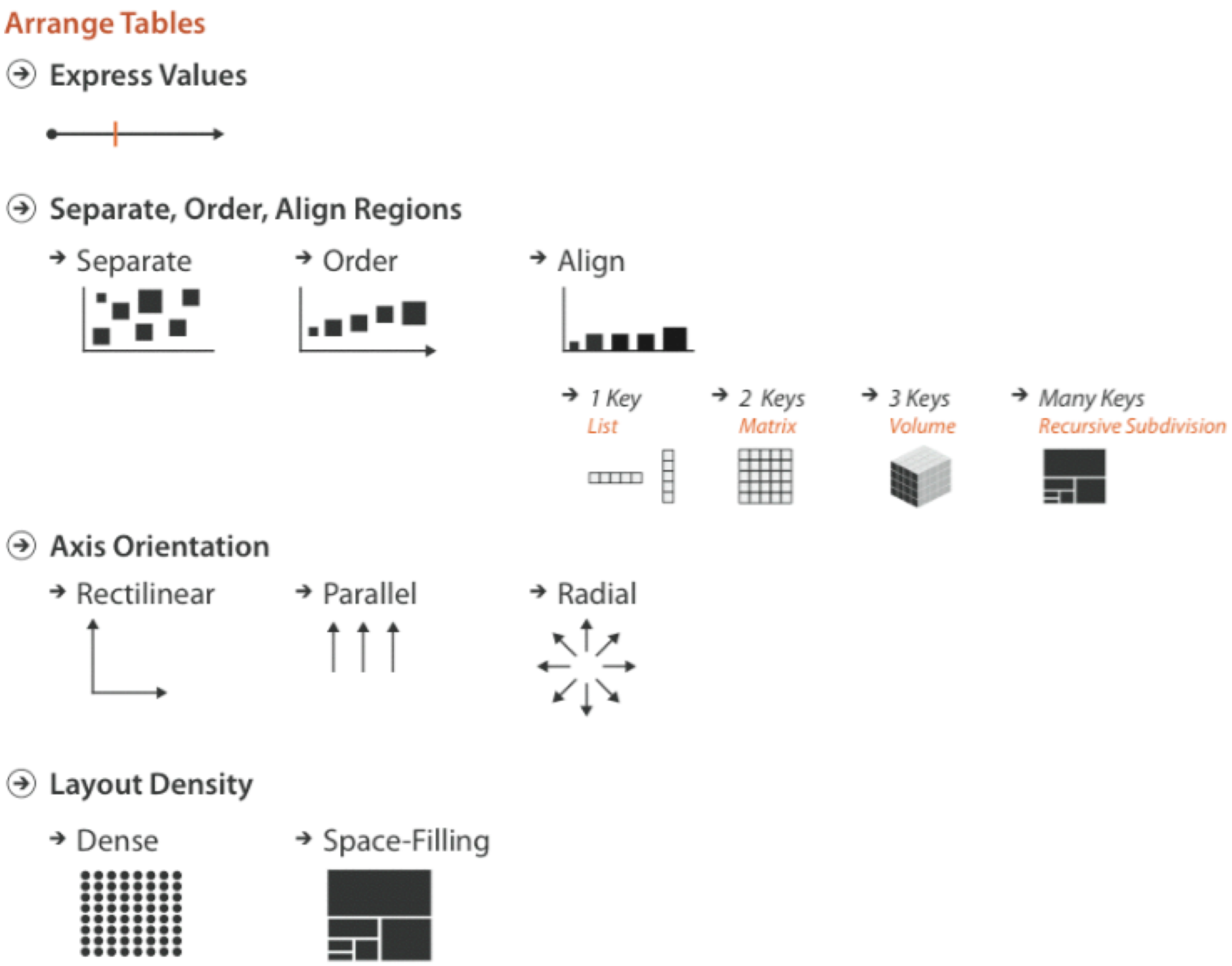
Reminders

2/26/2018 – Project 4 Peer Reviews Due

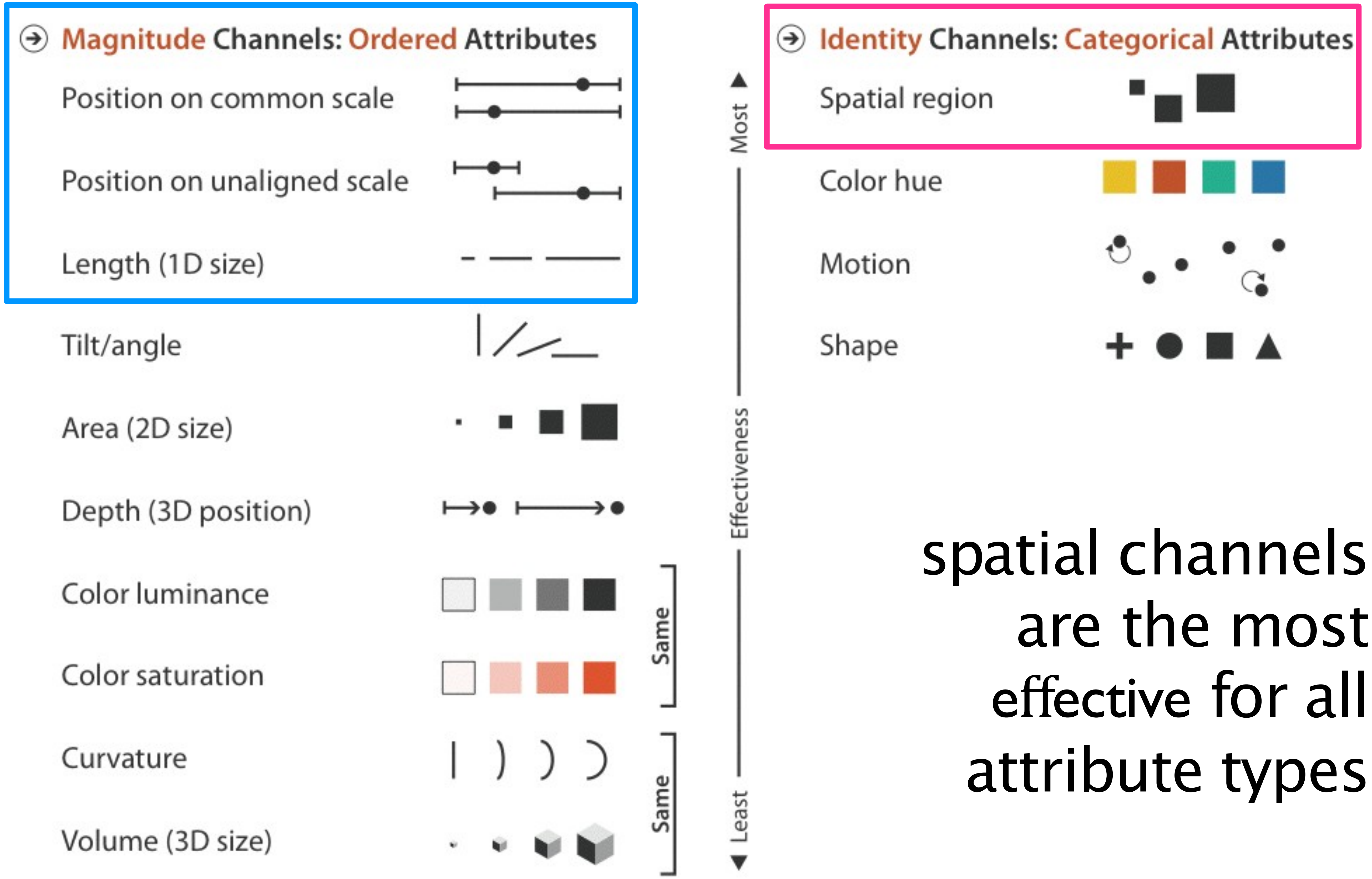
2/28/2018 – Project 5 Due

DATASET TYPES




Arrange Tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling

<u>**ARRANGE** IS THE FOCUS OF ALL FOUR DESIGN CHOICES FOR TABULAR DATA</u>

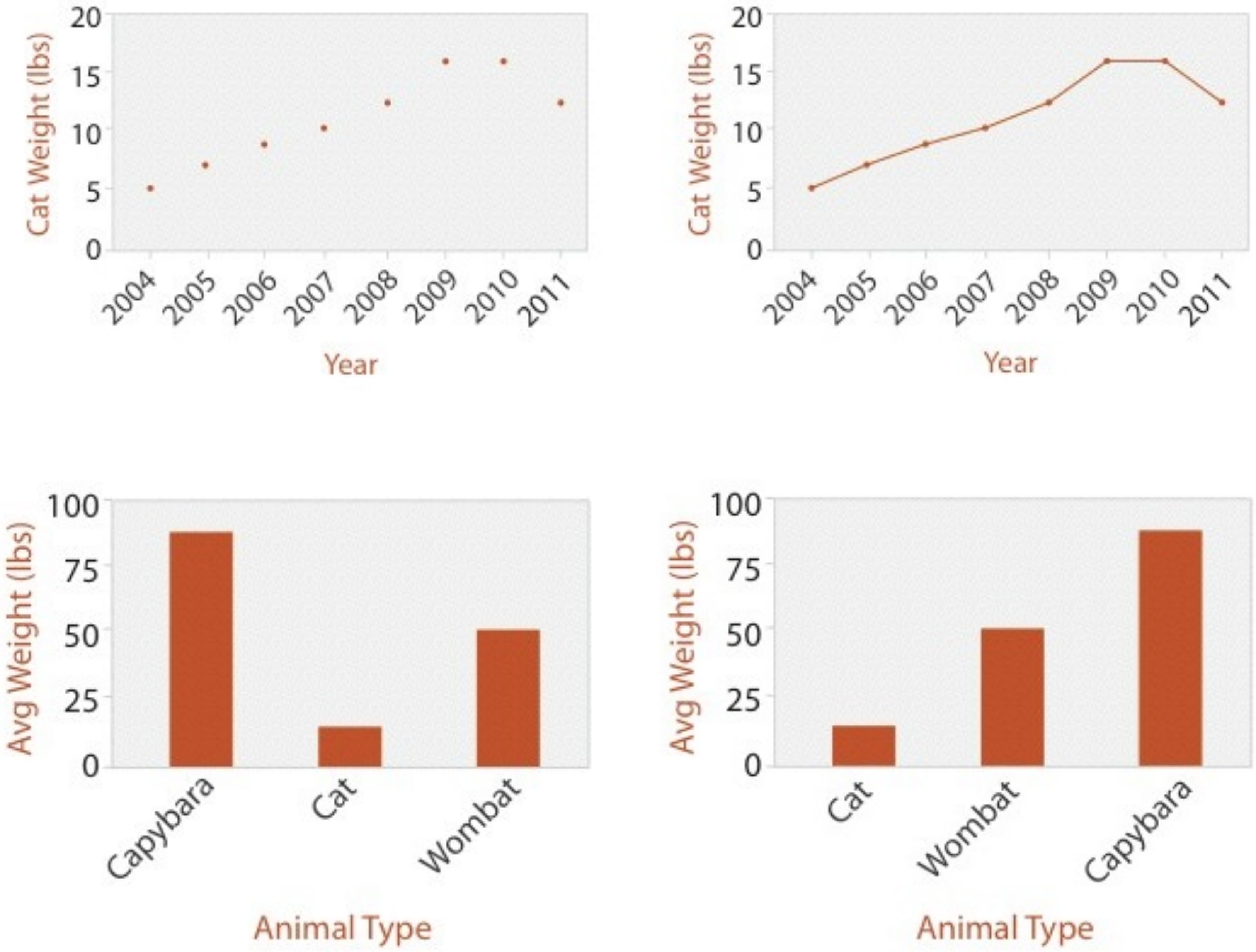


spatial channels are the most effective for all attribute types

Single Key, Single Value

ENCODE ONE KEY ATTRIBUTE

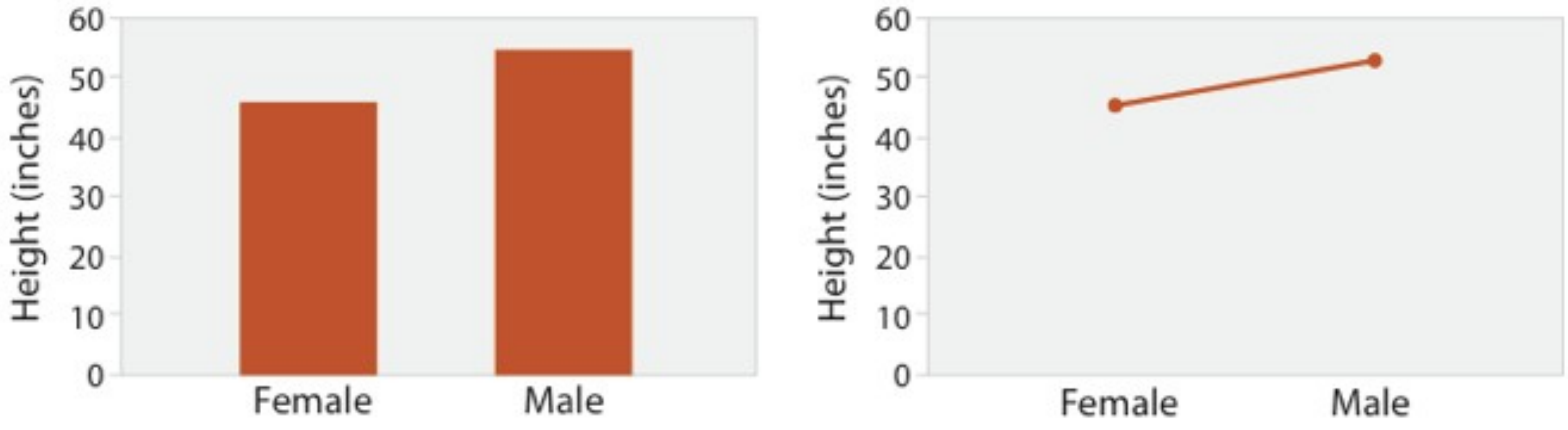
BAR, DOT, & LINE CHARTS



DON'T USE LINE CHARTS FOR CATEGORICAL ATTRIBUTES!

ok: "Men are taller than women (on average)"

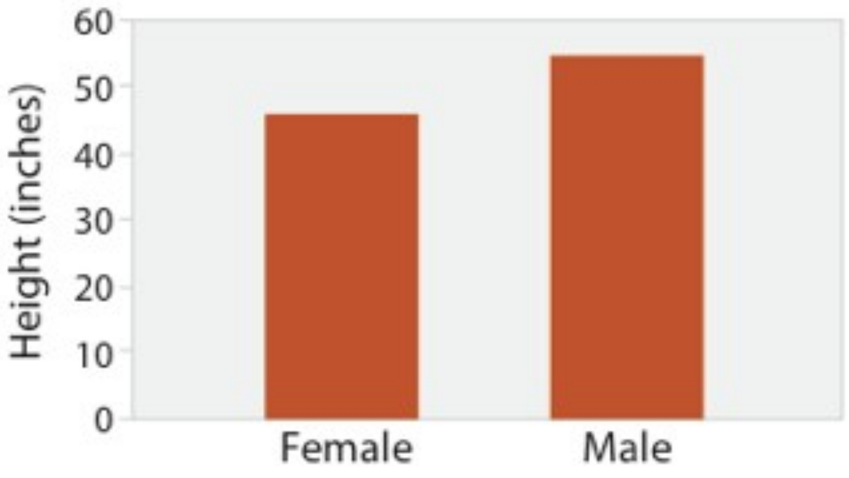
bad: "The more male a person is, the taller he/she is"

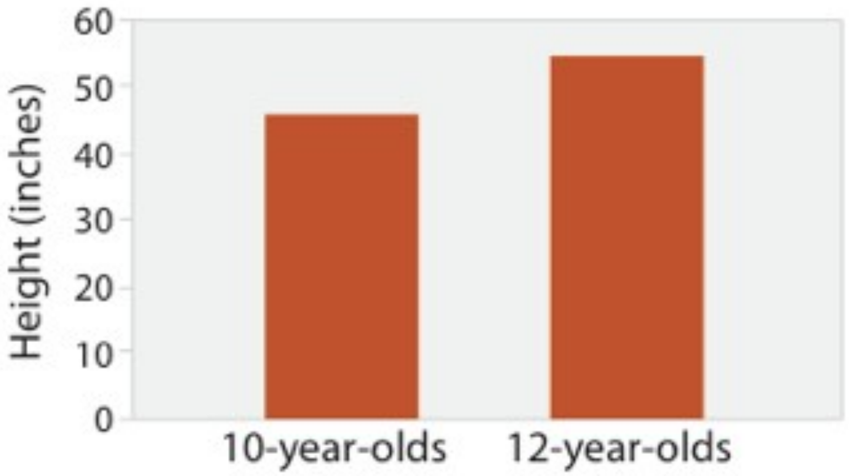


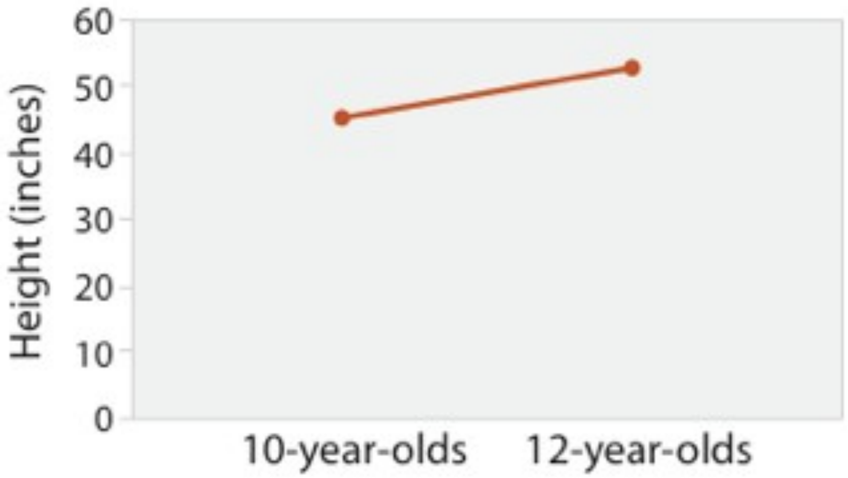
DON'T USE LINE CHARTS FOR CATEGORICAL ATTRIBUTES!

ok: "Men are taller than women (on average)"

bad: "The more male a person is, the taller he/she is"







ok: "Twelve year olds are taller than ten year olds"

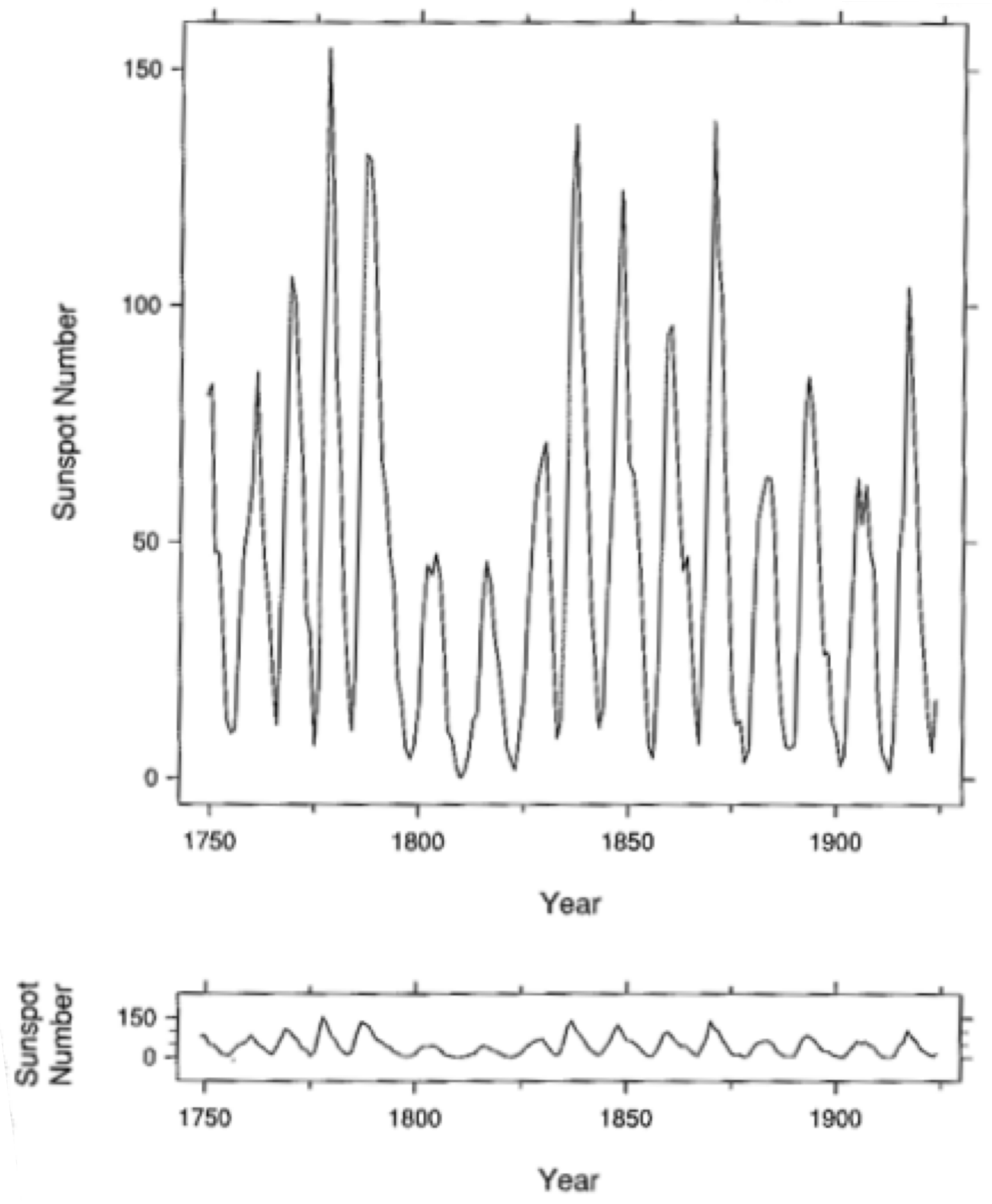
ok: "Height increases with age"

CLEVELAND 1994

BANK TO 45°

The aspect ratio of a graph is an important factor for judging rate of change.

perceptual principle: most accurate angle judgment is at 45°

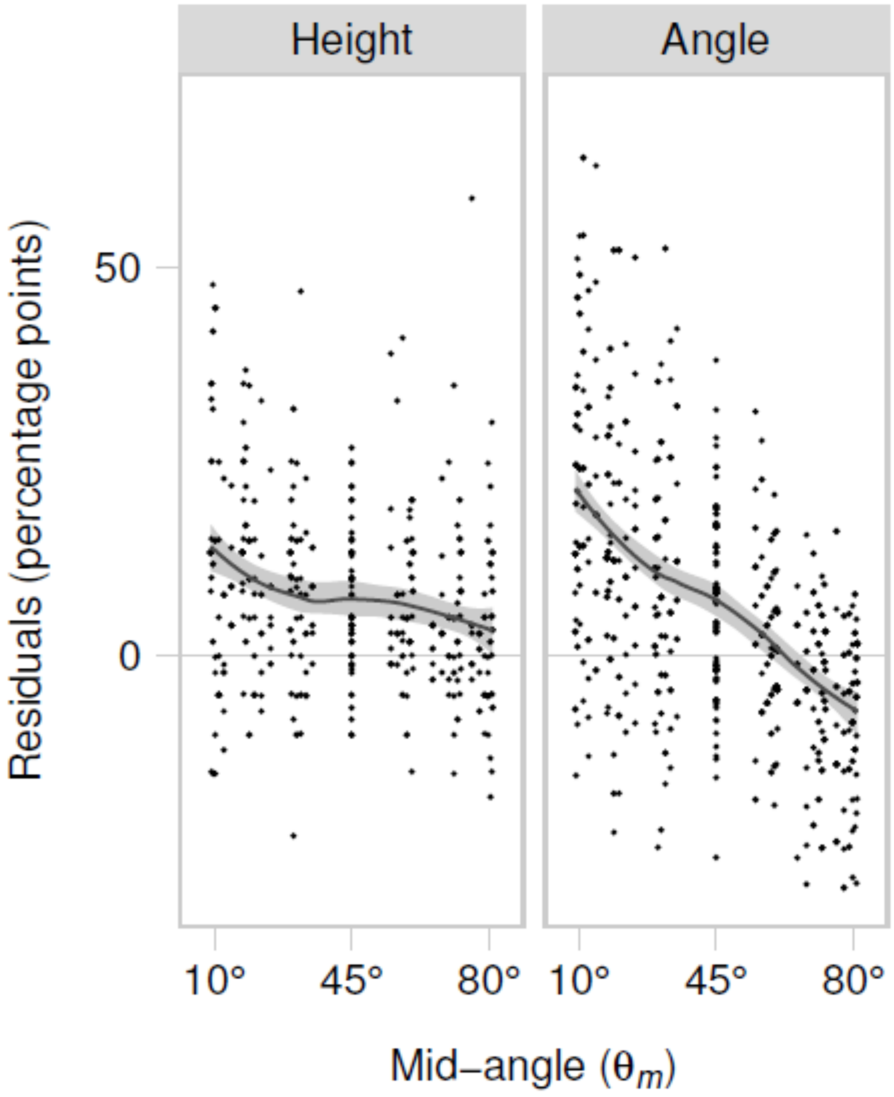


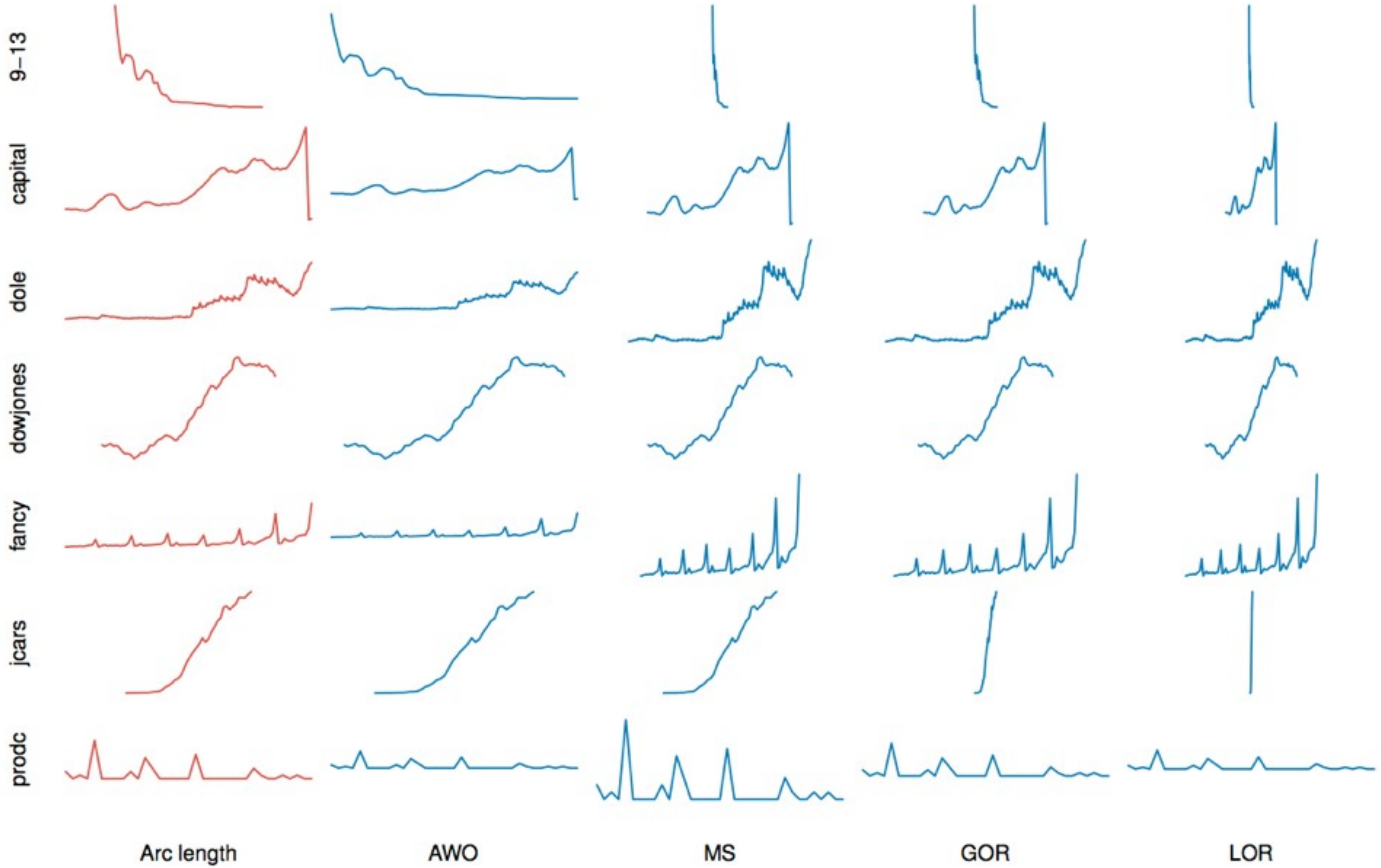
COUNTER-POINT

TALBOT 2012

people use two different strategies to estimate slope—angle and height

slope angle accuracy NOT minimized at 45°

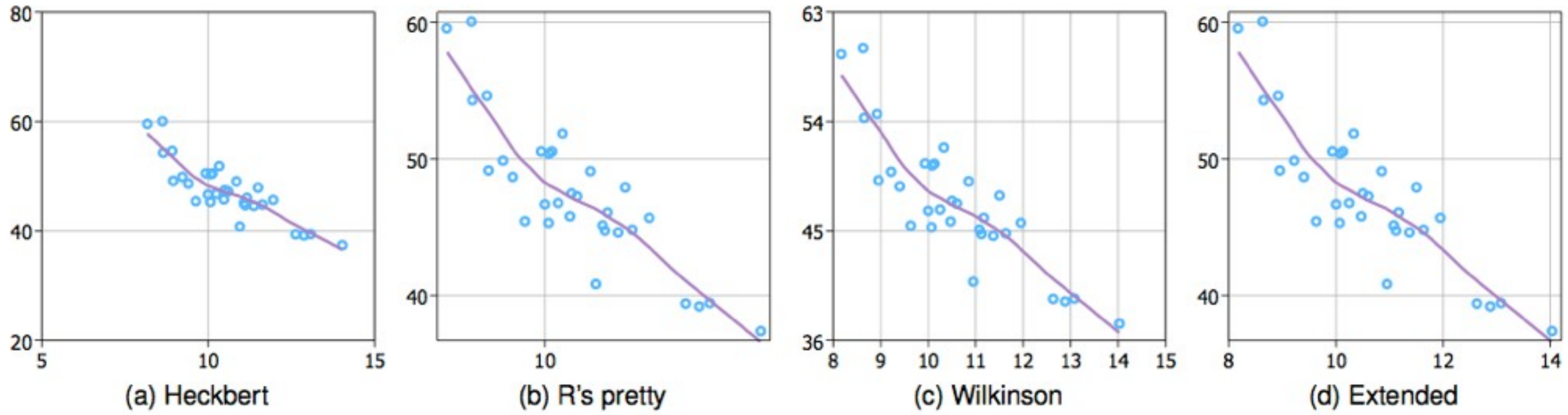



9–13
capital
dole
dowjones
fancy
jcars
prodc
Arc length
AWO
MS
GOR
LOR

TALBOT 2011

Tick Placement

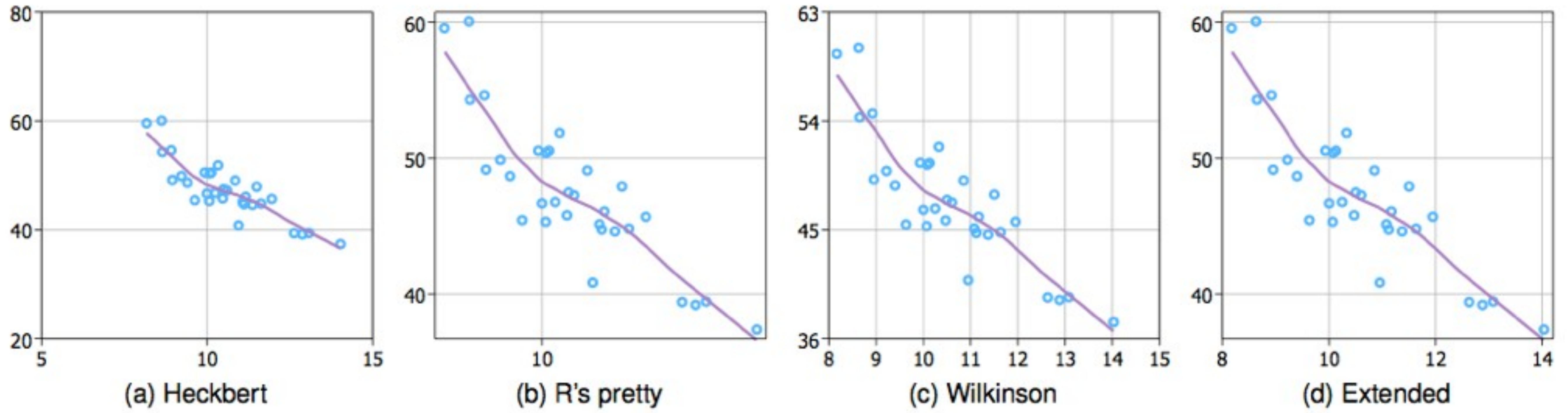
Ticks help in user interpretation of data, but too much may hinder



AUTOMATIC TICKS

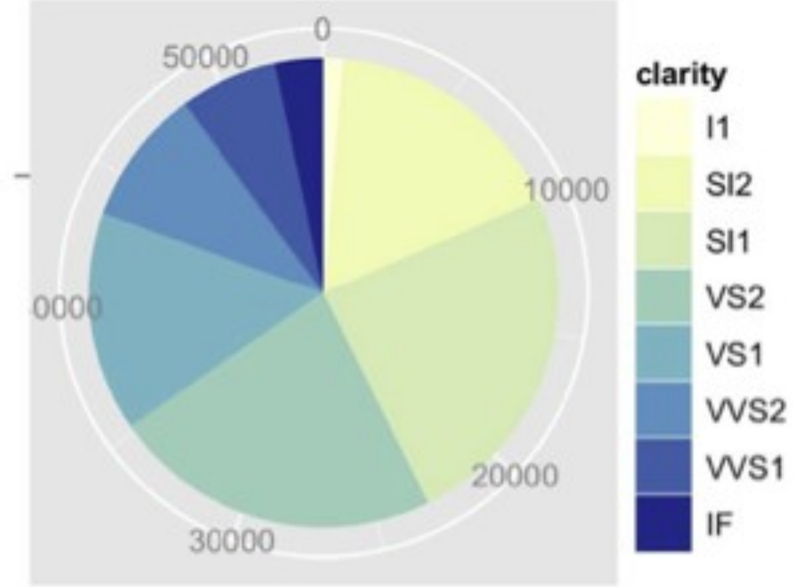
optimization of label formatting, font size, and orientation

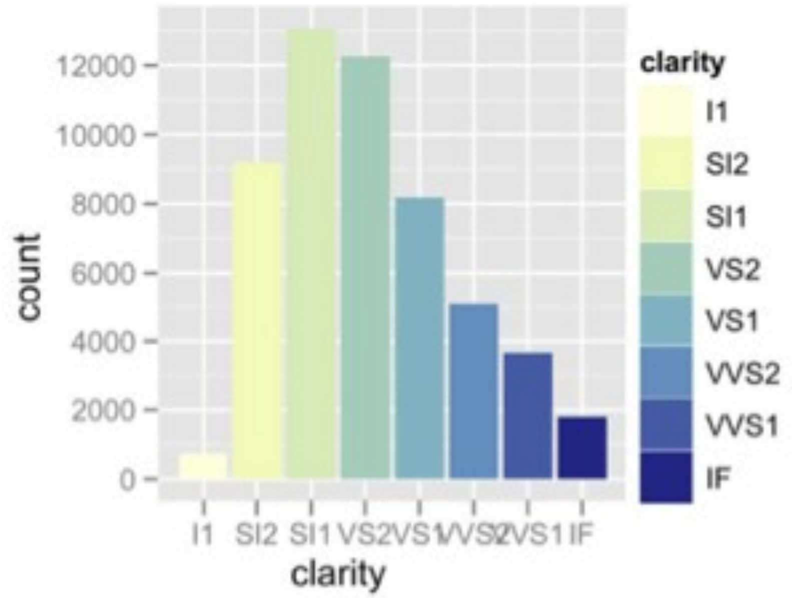
placement based on simplicity, coverage, granularity, and legibility

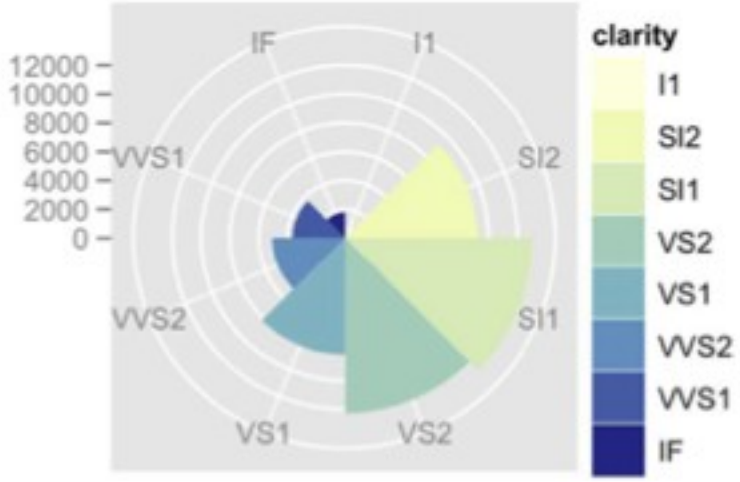


TALBOT 2010

PIE CHARTS: TAKE CARE WITH ACCURACY







2 KEYS, 1 VALUE

ENCODE USING TWO KEYS: HEATMAP

uses heatmap representation

matrix layout using keys

encode values with color

often augmented with clustering

ENCODE USING TWO KEYS: HEATMAP

uses heatmap representation

matrix layout using keys

encode values with color

often augmented with clustering

0.2	0.4	1	1	1	0.8
1	0	0	0	1	1
0.7	0.8	1	1	0.8	0.6
1	0	0.2	0.5	1	1
0.5	0.8	0.5	0.3	0.5	0.8
0.7	0.5	0.8	0.7	1	1
1	0.3	0.4	1	1	1
0.5	0	0	0.7	0.5	0.3

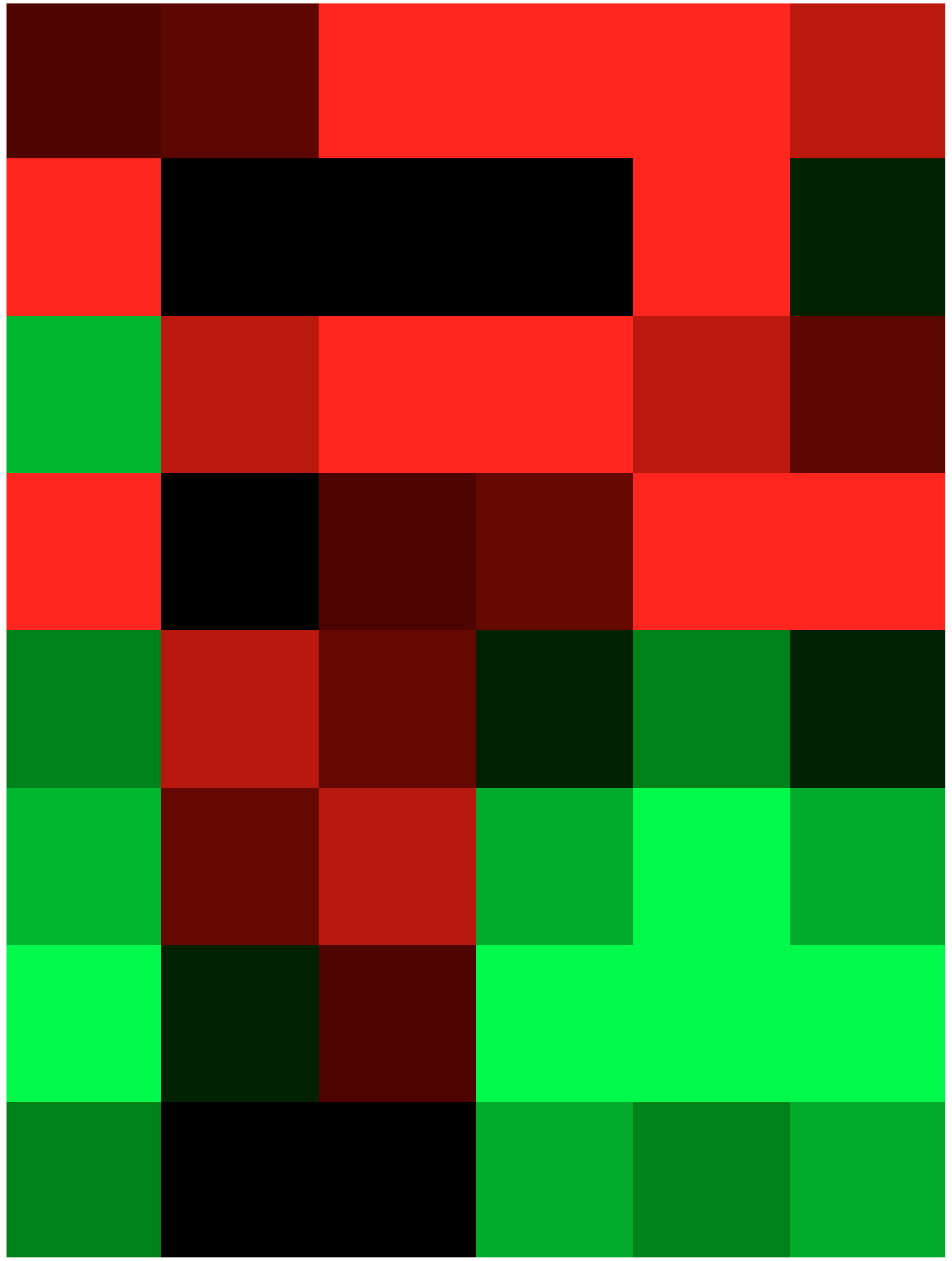
ENCODE USING TWO KEYS:

HEATMAP

uses heatmap representation

- matrix layout using keys
- encode values with color

often augmented with clustering



ENCODE USING TWO KEYS: HEATMAP

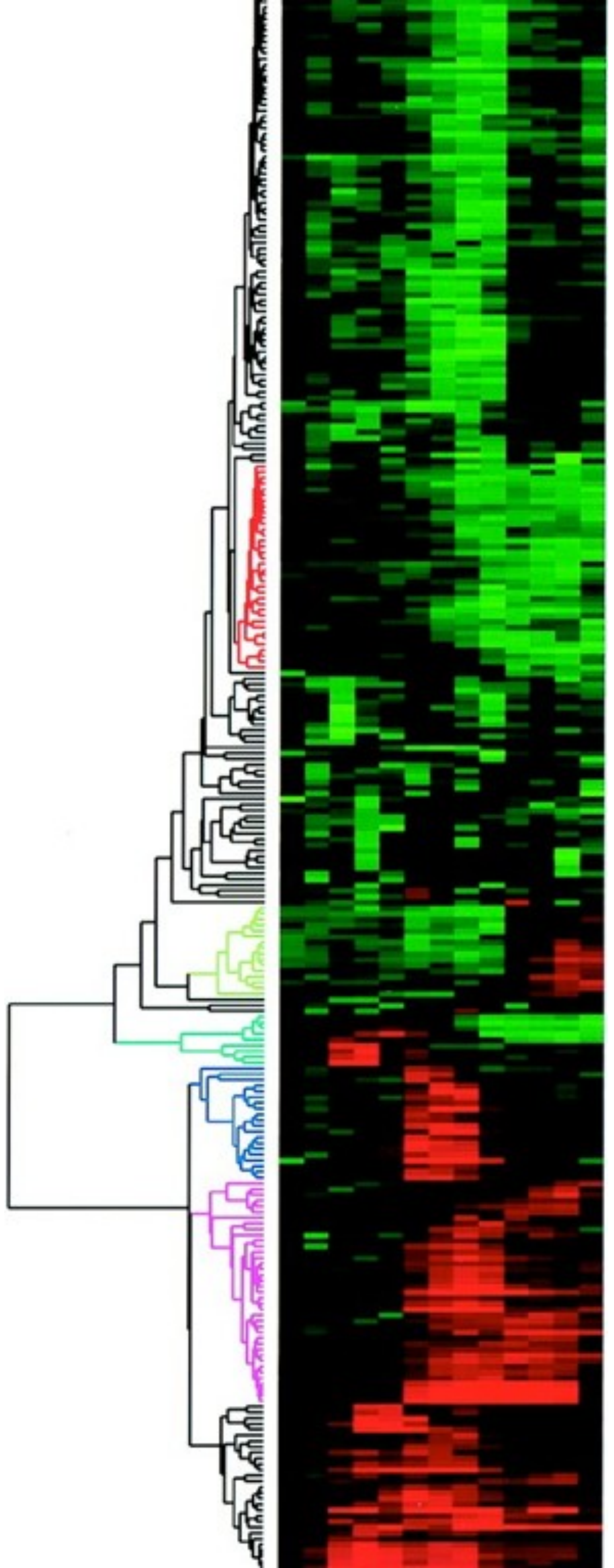
uses heatmap representation

matrix layout using keys

encode values with color

often augmented with clustering

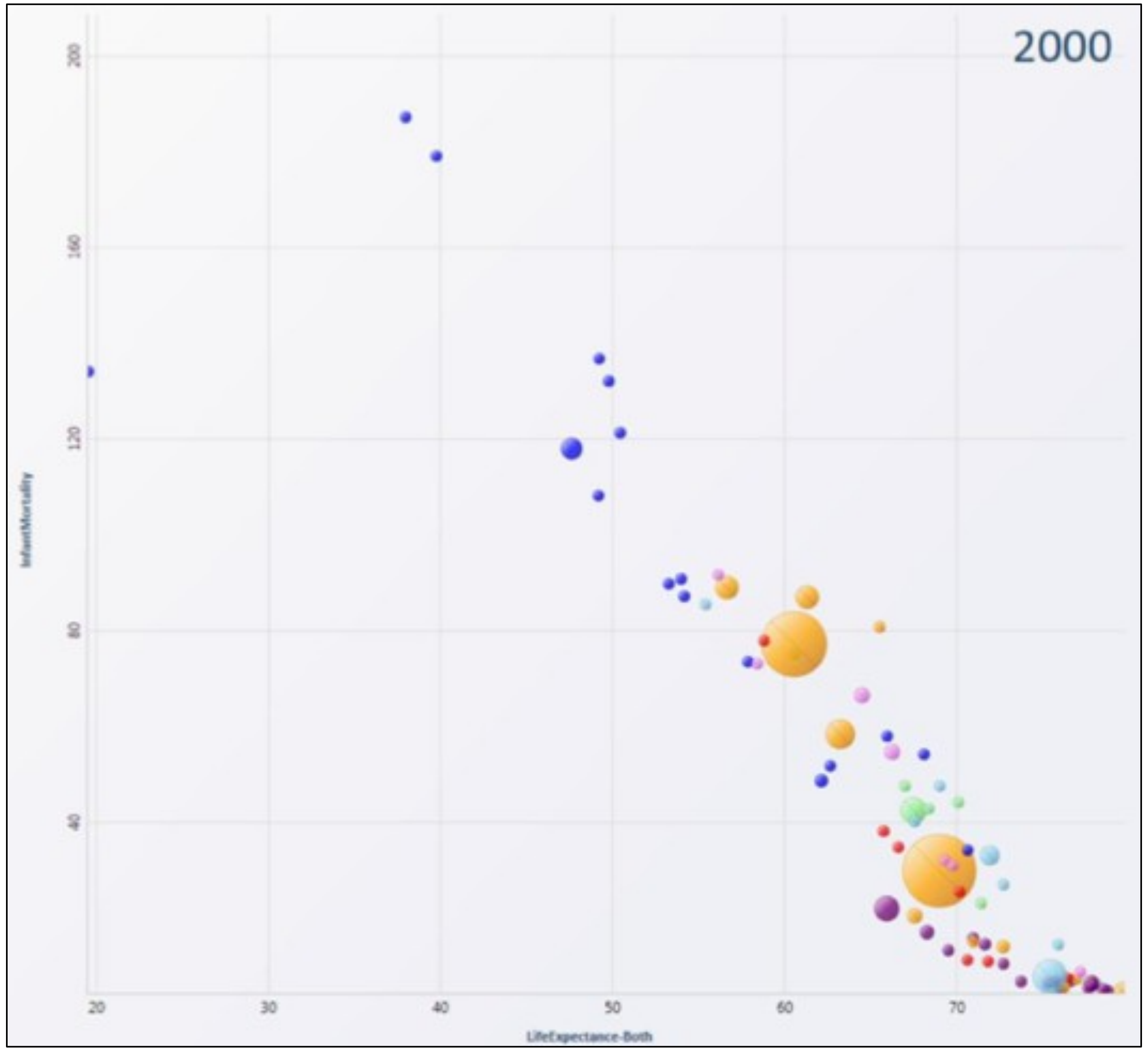
here, used on genomic data



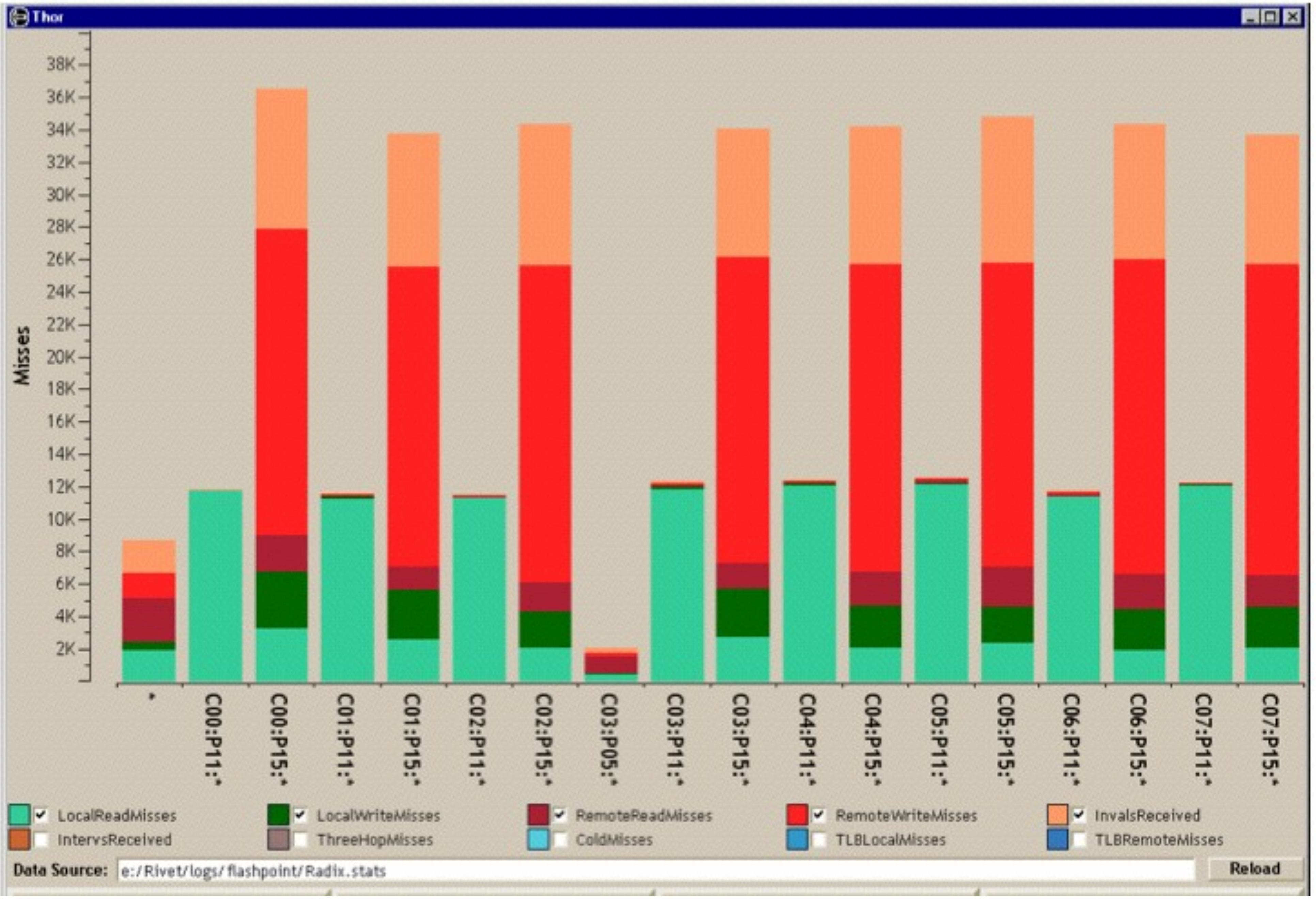
EISEN 1998

MULTIPLE ATTRIBUTES

ENCODE USING SCATTERPLOTS



ENCODE USING STACKED BAR CHART



ALIGN USING MULTIPLE KEYS

LineUp: Visual Analysis of Multi-Attribute Rankings

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit

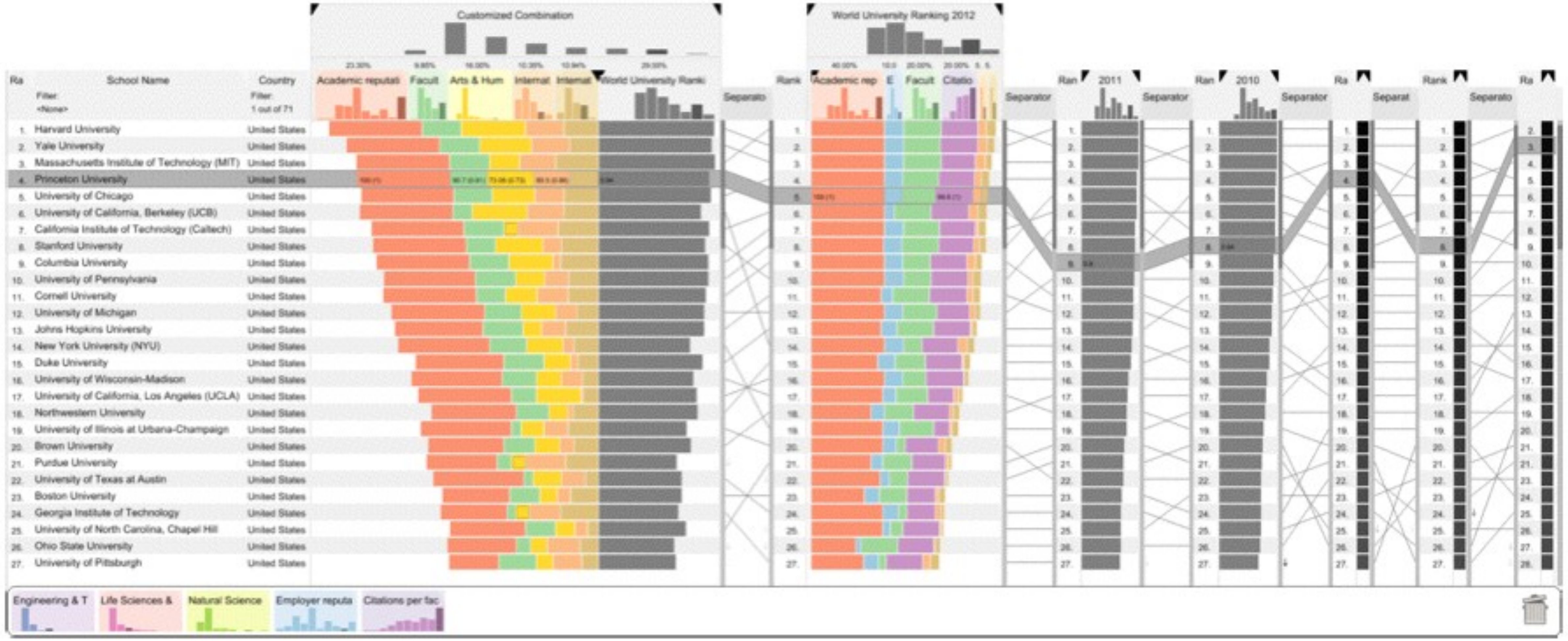


Fig. 1. LineUp showing a ranking of the top Universities according to the QS World University Ranking 2012 dataset with custom attributes and weights, compared to the official ranking.

Abstract— Rankings are a popular and universal approach to structuring otherwise unorganized collections of items by computing a rank for each item based on the value of one or more of its attributes. This allows us, for example, to prioritize tasks or to evaluate the

Challenge

rankings based on single attribute are trivial to display

when based on multiple attributes:

not clear how attributes contribute to ranking

not clear how changes to multiple attributes will affect ranking

different contexts/people/situations will rank on multiple attributes differently

LineUp
Visual Analysis of Multi-Attribute Rankings

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit

CALEYDO

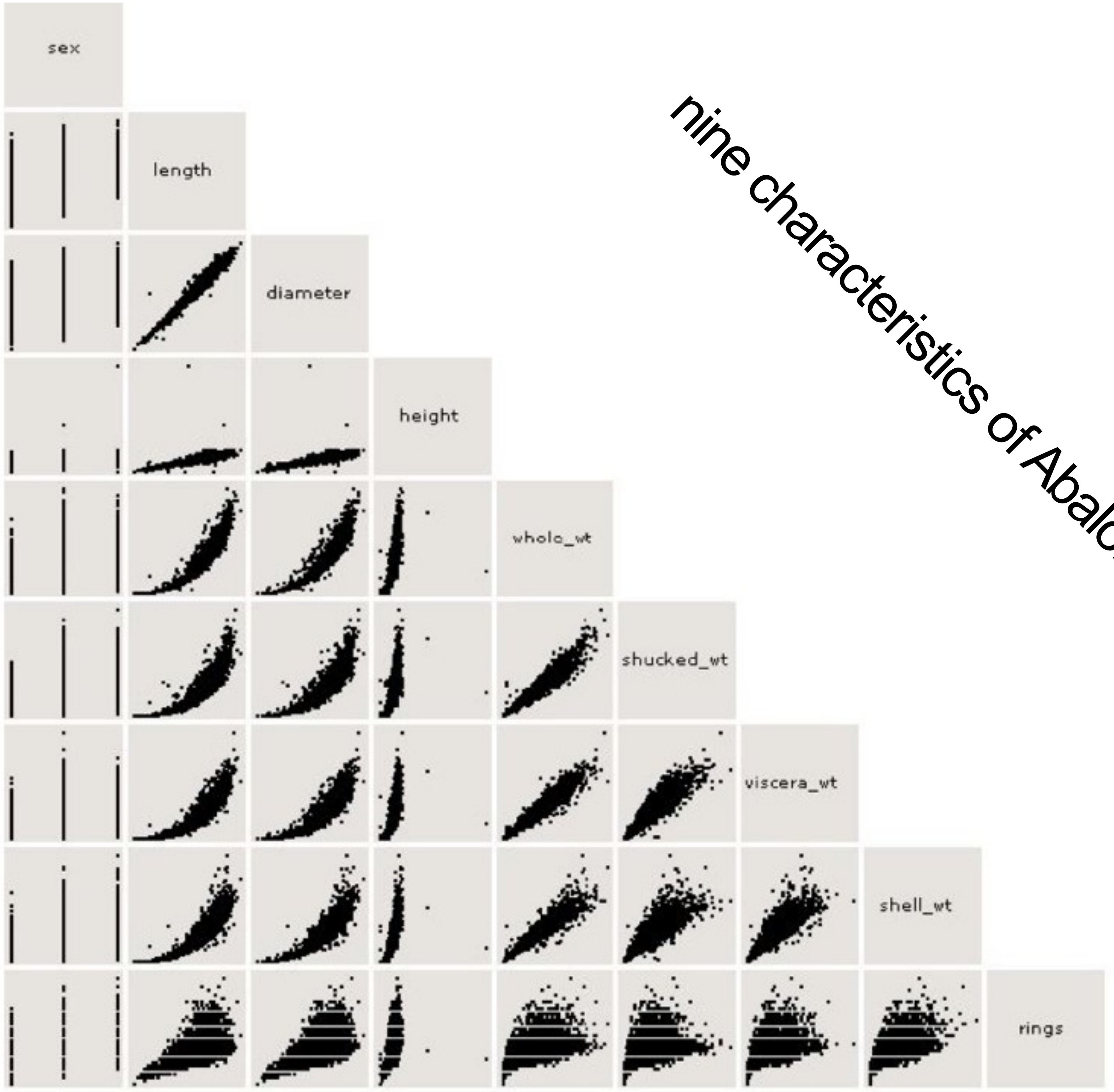
JKU
JOHANNES KEPLER
UNIVERSITÄT LINZ

HARVARD
School of Engineering
and Applied Sciences

HARVARD
MEDICAL SCHOOL

CRITIQUE: WHAT DO YOU THINK?

SPLOMs: SCATTERPLOT MATRICES



nine characteristics of Abalone (sea snails)

PARALLEL COORDINATES

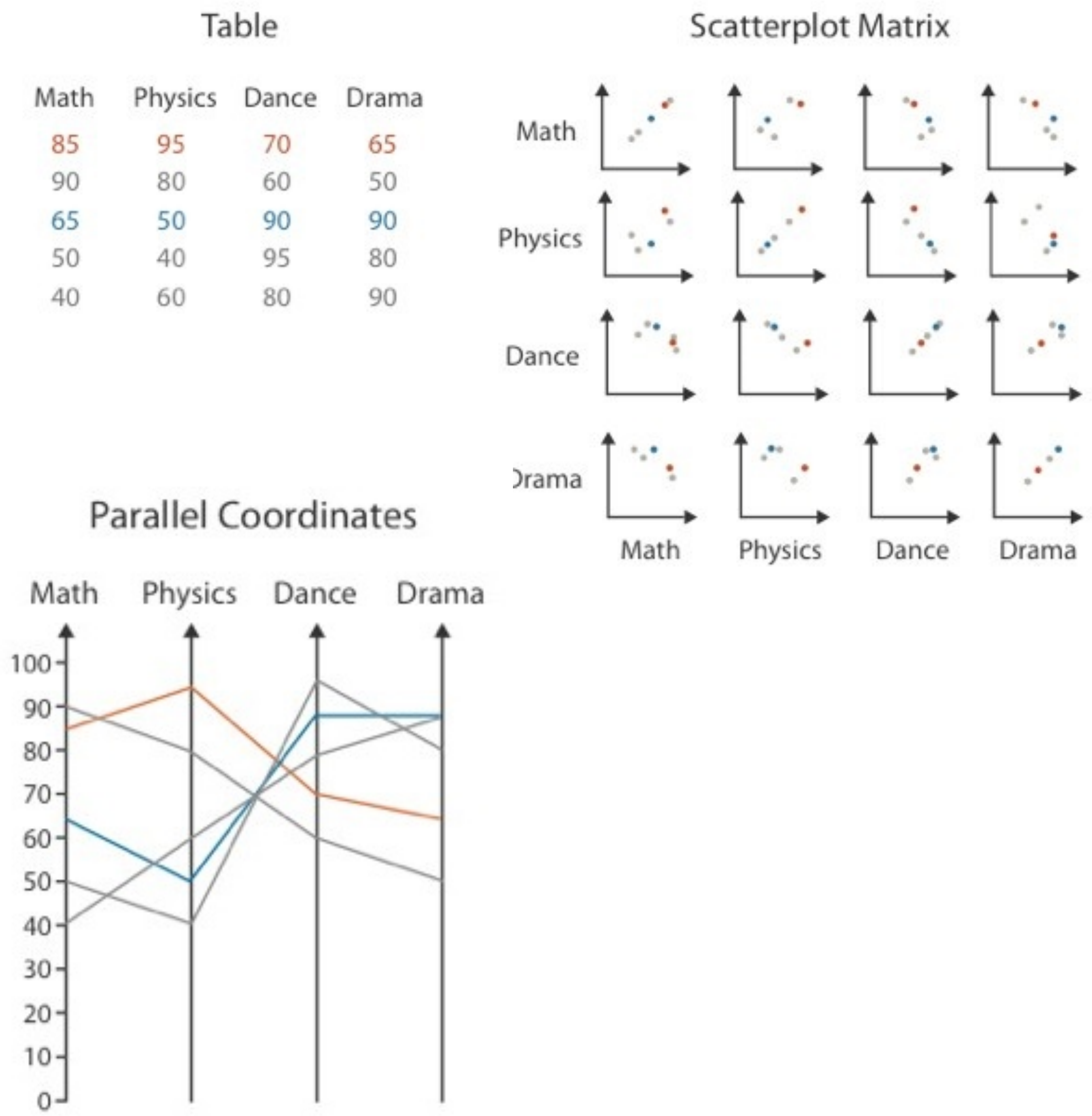
scatterplot limitation: visual representation with orthogonal axes

can show only two attributes with spatial position channel

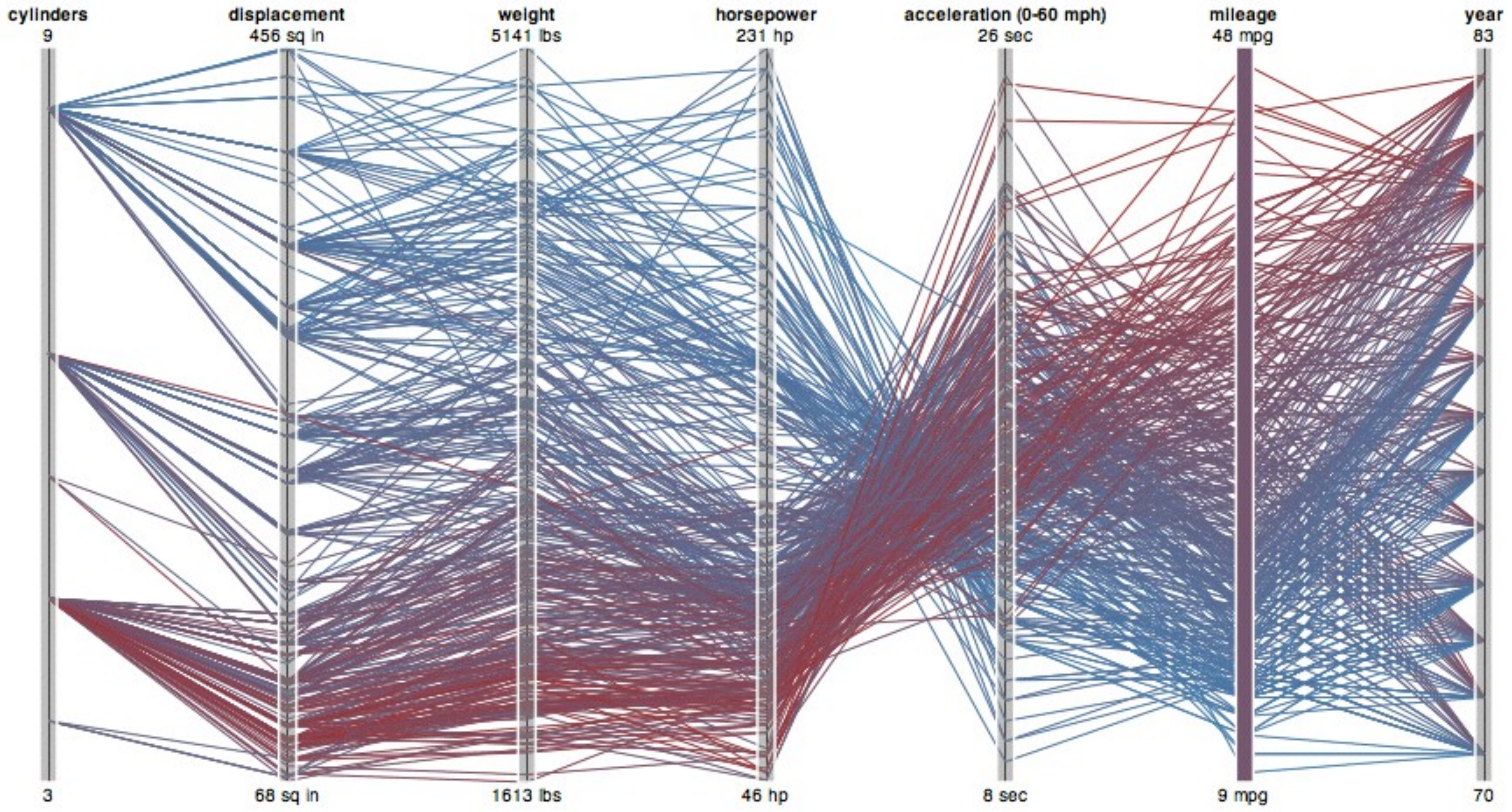
alternative: line up axes in parallel to show many attributes with position

item encoded with a line with n segments

n is the number of attributes shown



PARALLEL COORDINATES

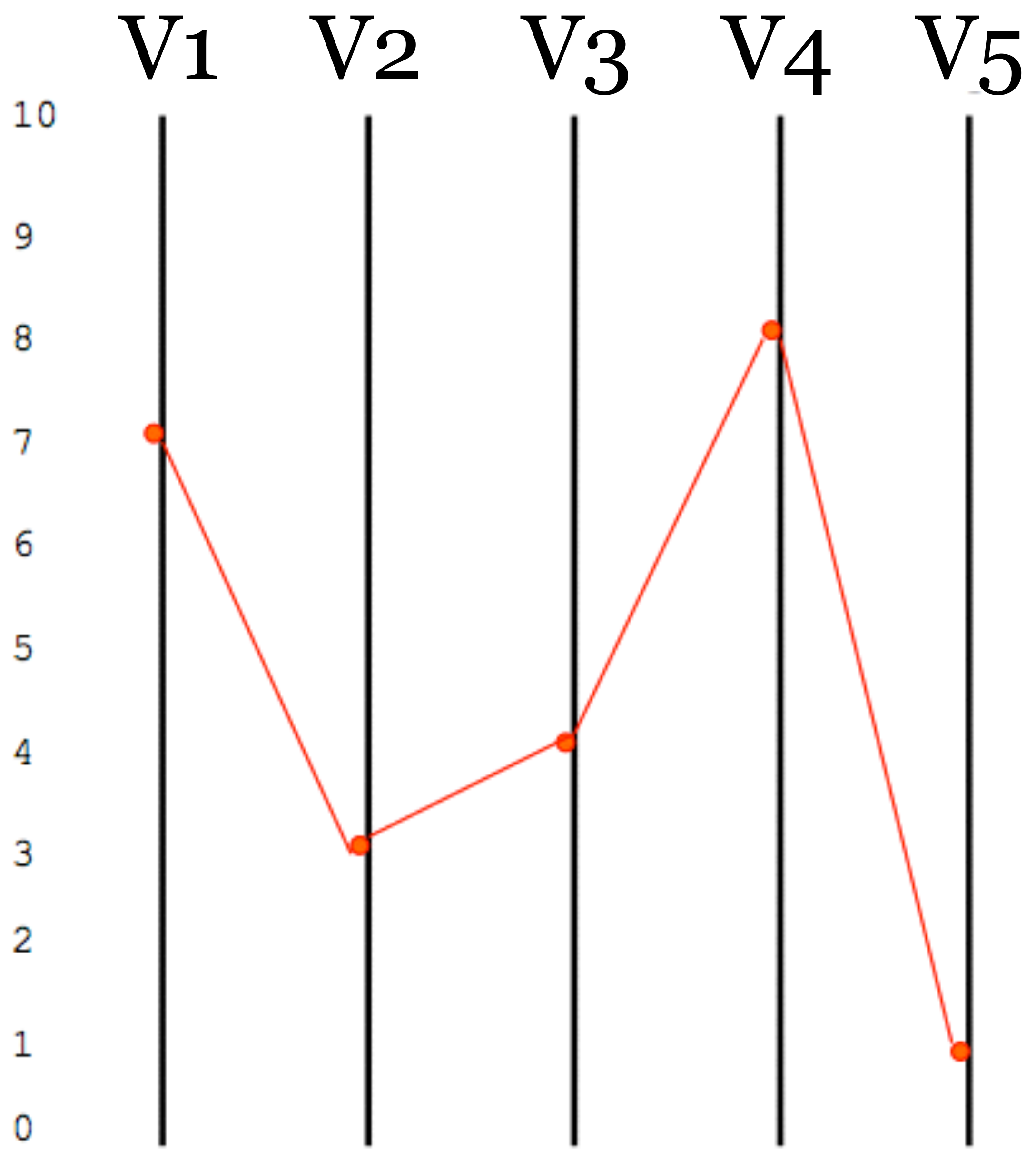


PROTOVIS

EXAMPLE

	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

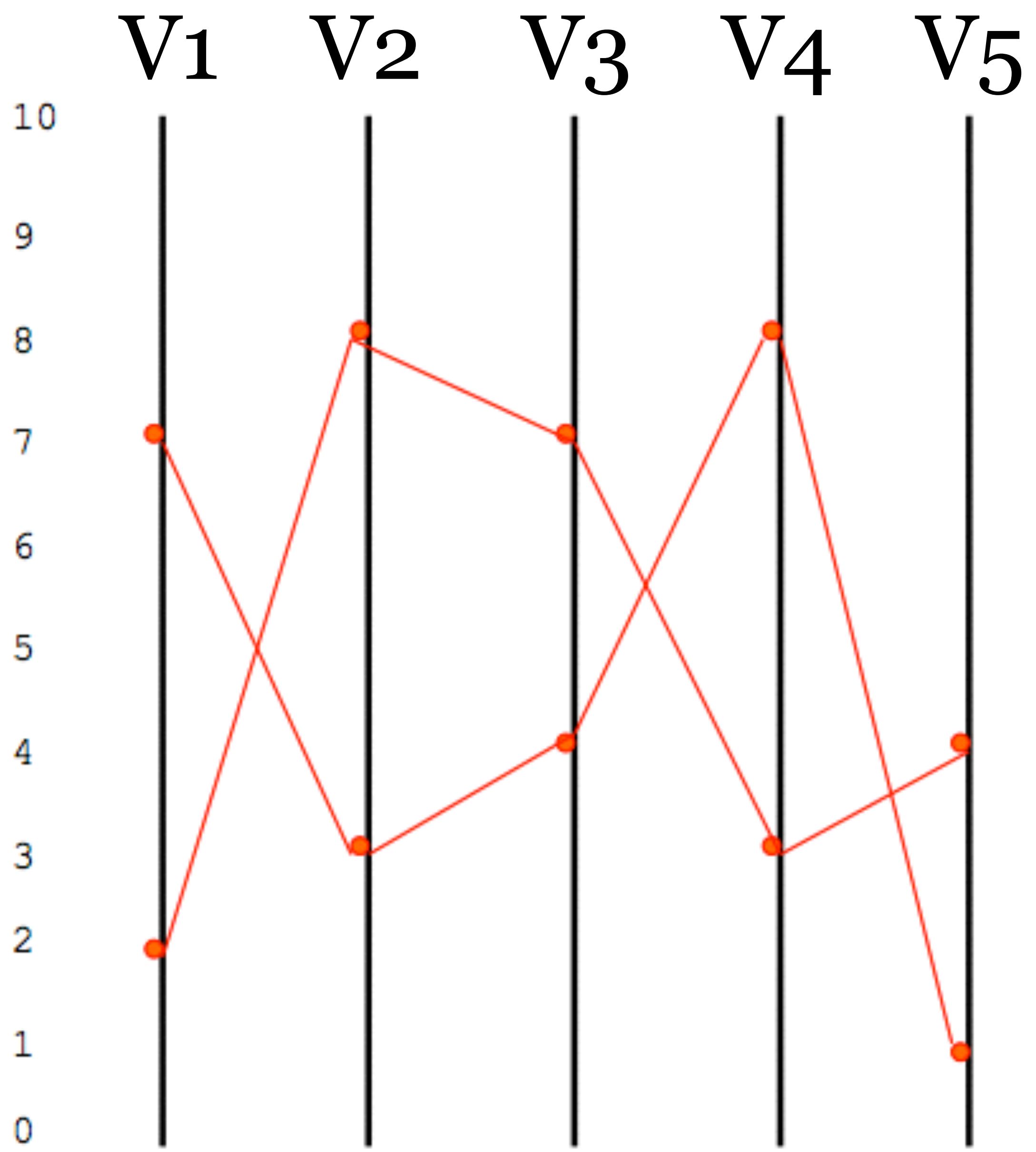
EXAMPLE



D1

7 3 4 8 1

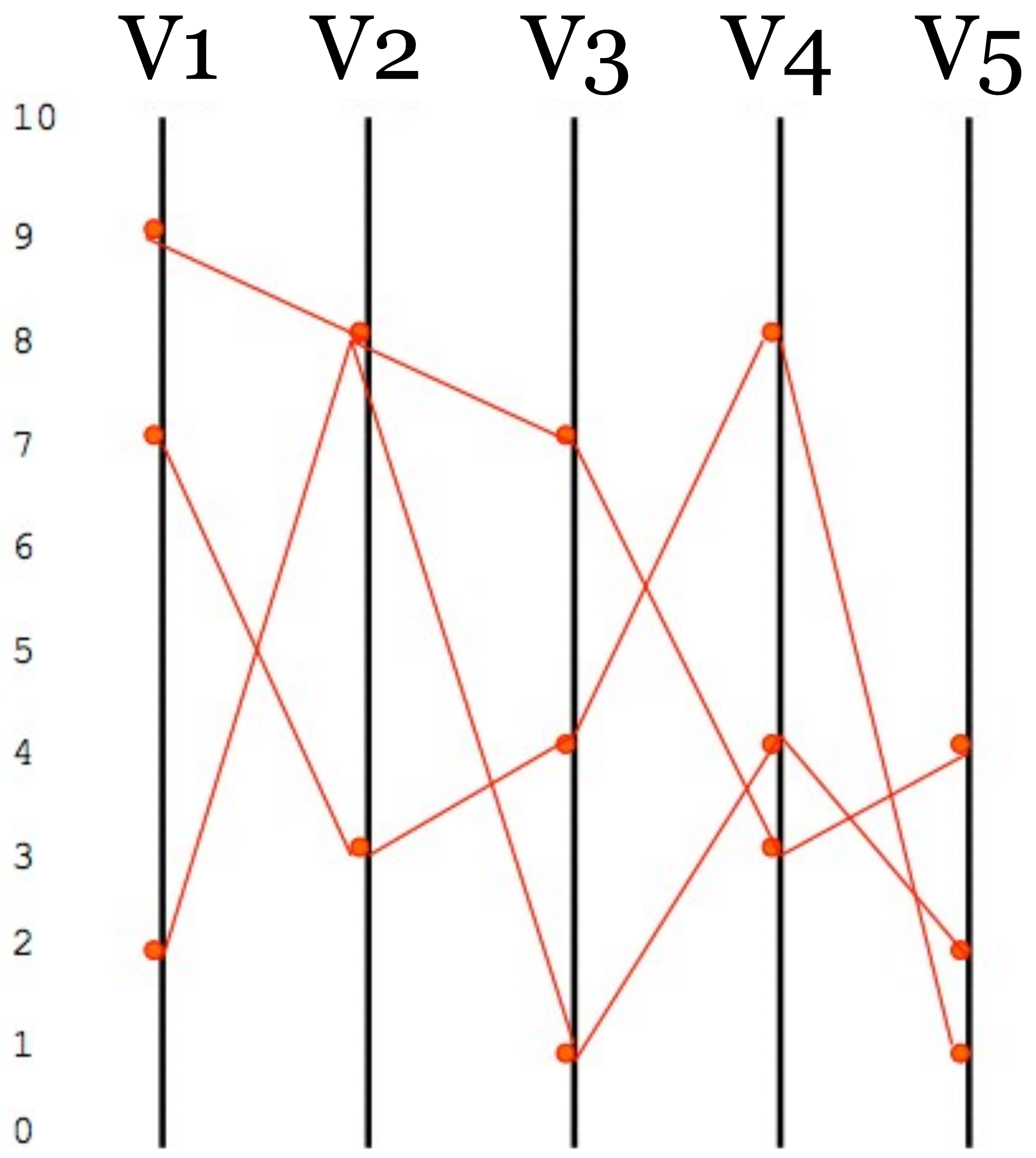
EXAMPLE



D2

2 7 6 3 4

EXAMPLE



D3

9 8 1 4 2

PARALLEL COORDINATES TASK

show correlation

positive correlation: straight lines

negative correlation: lines cross at a single point

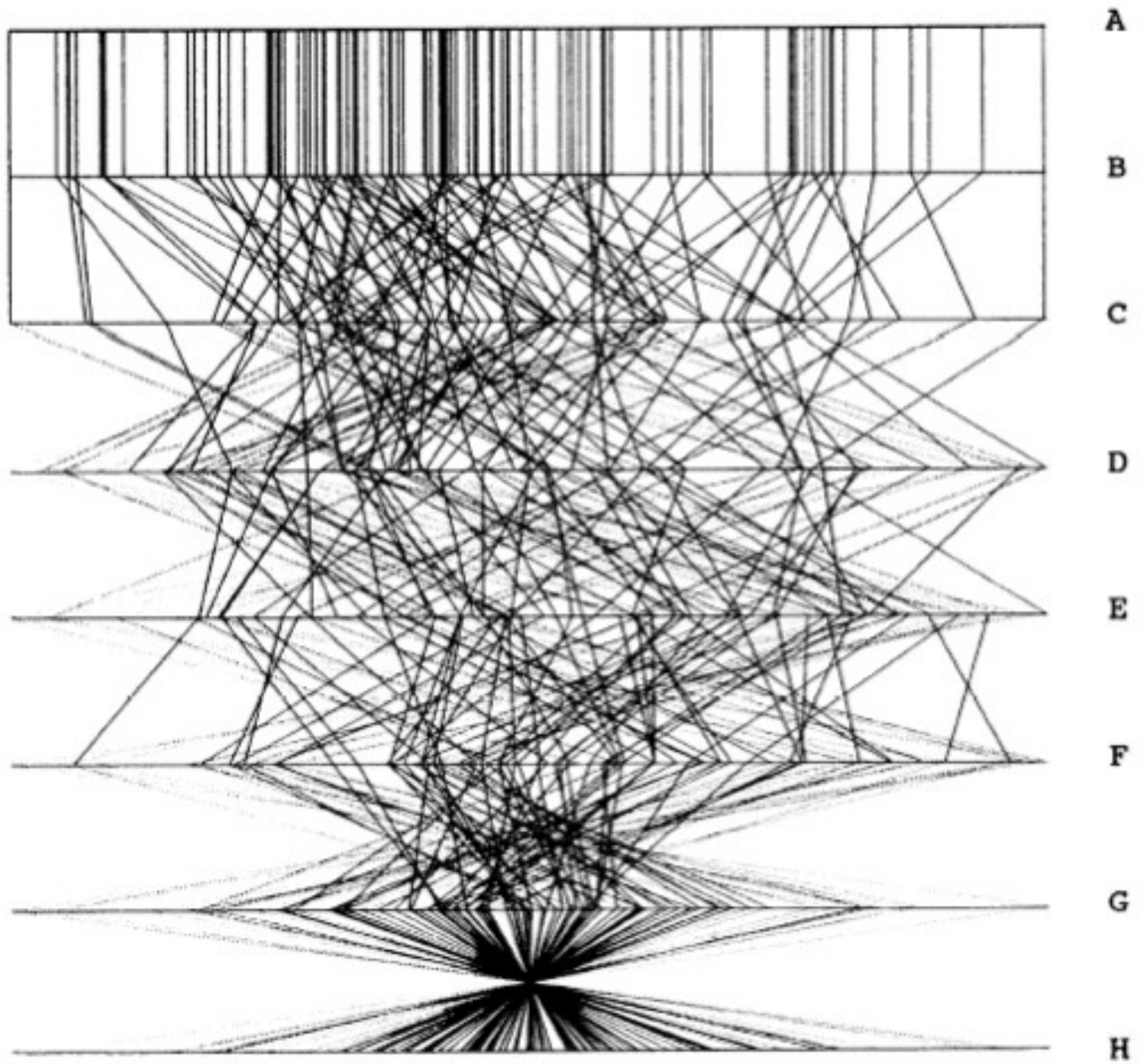
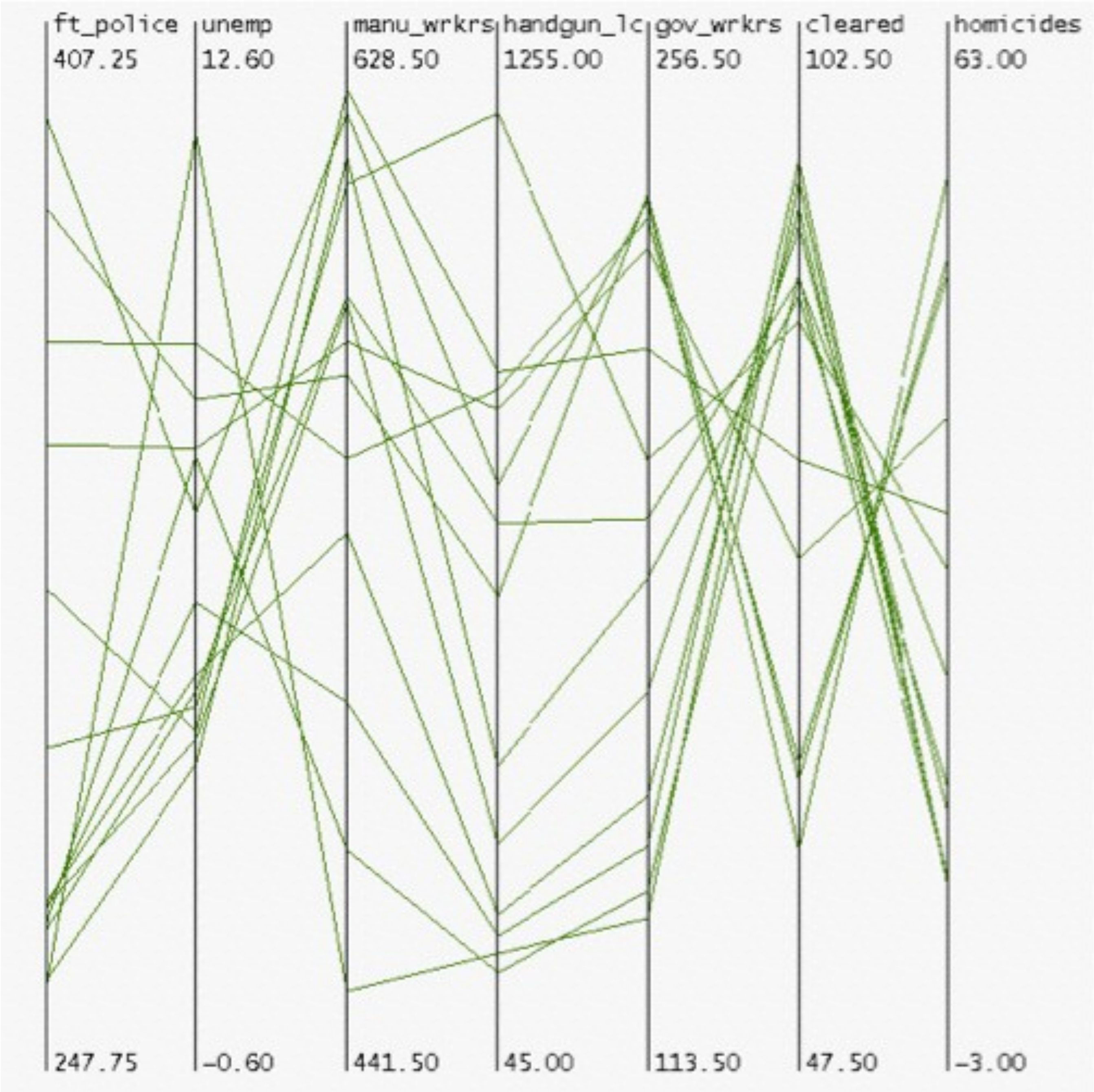


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of ρ = 1, .8, .2, 0, −.2, −.8, and −1.

WEGMAN 1990

PARALLEL COORDINATES TASK

do you see any correlations?



FUA 1999

PARALLEL COORDINATES TASK

visible patterns only between neighboring axis pairs

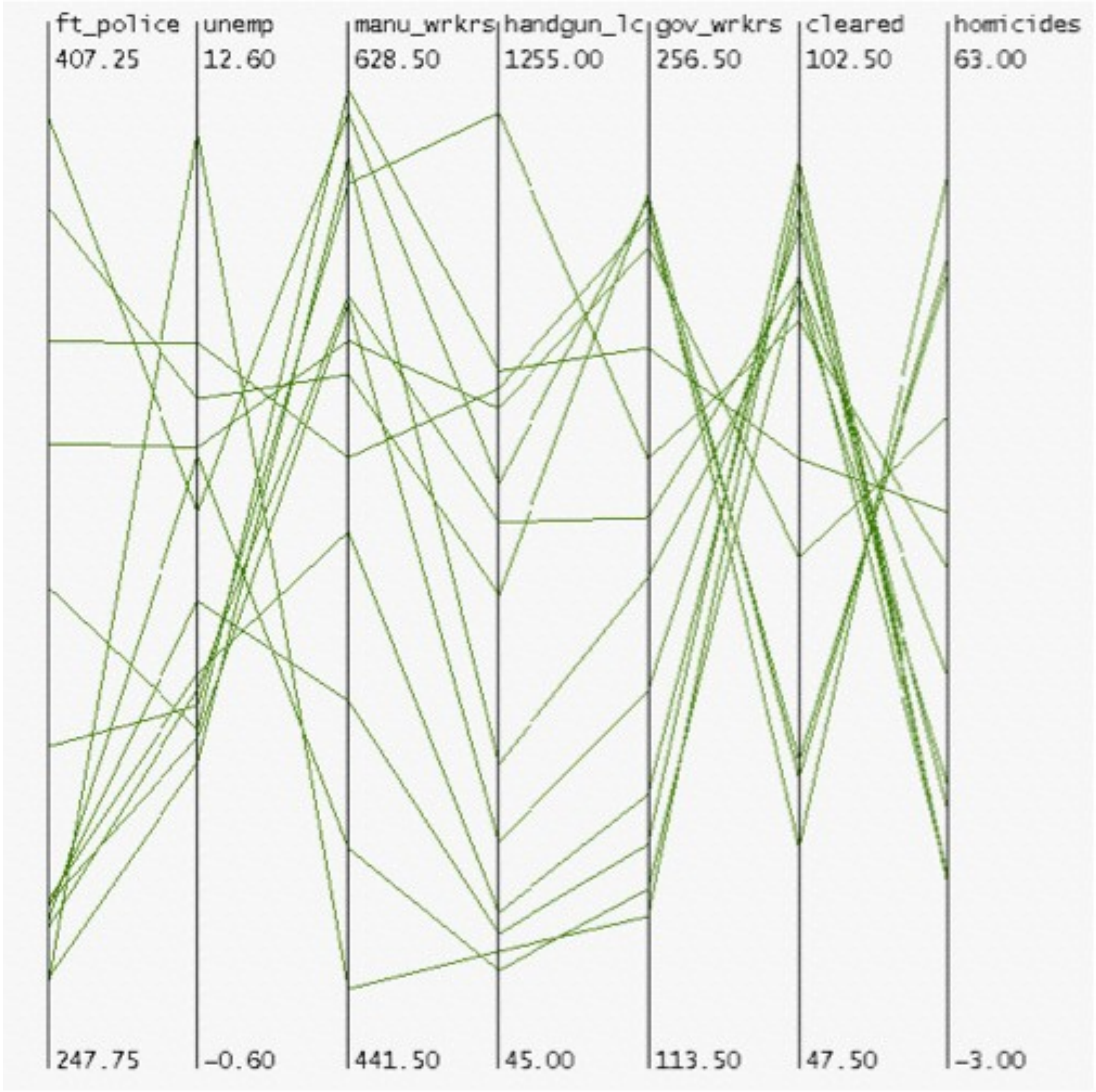
how to pick axis order?

usual solution: reorderable axes, interactive exploration

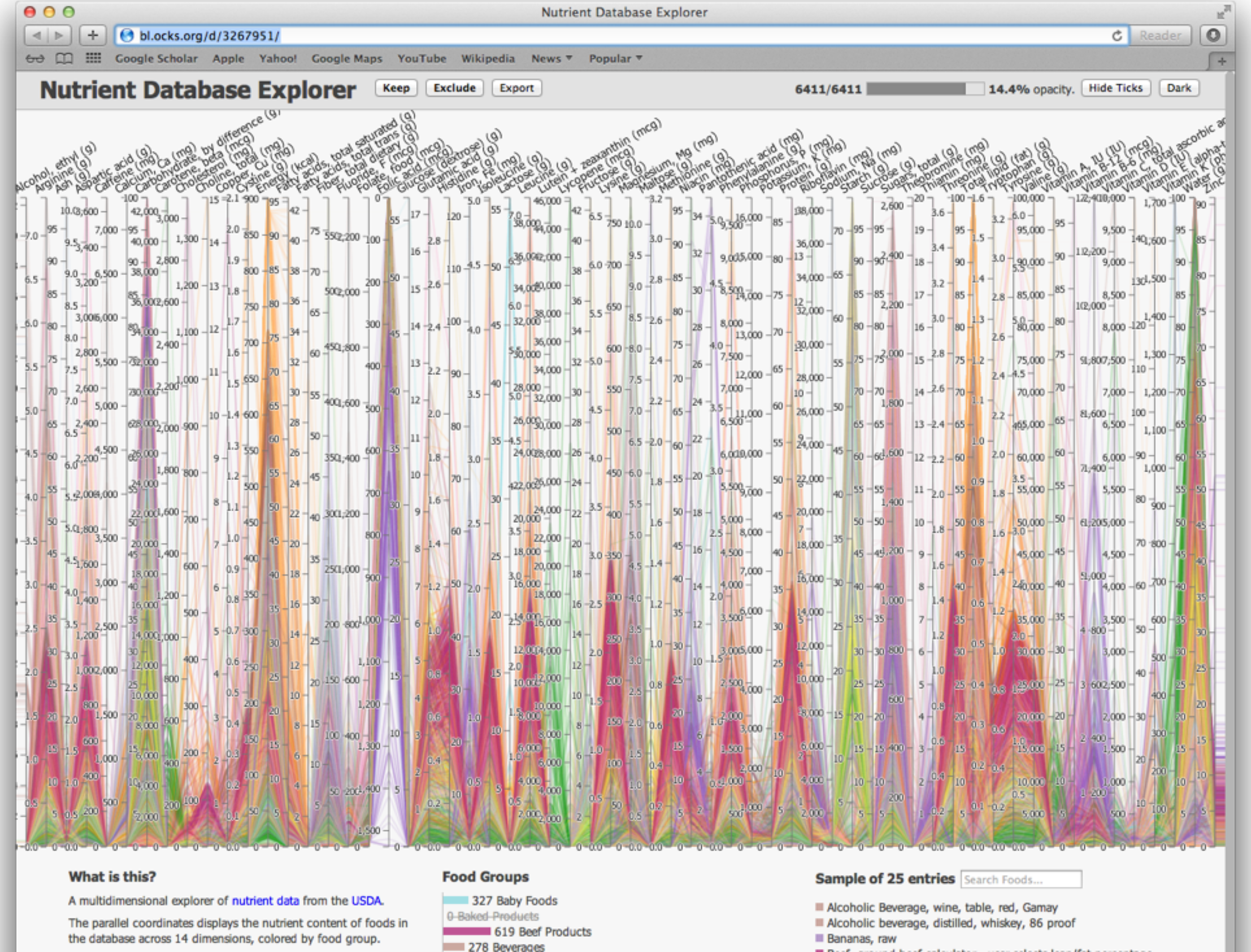
same weakness as many other techniques

downside: human-powered search

not directly addressed in HPC paper



FUA 1999



Nutrient Database Explorer
bl.ocks.org/d/3267951/
Google Scholar
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News
Popular
Nutrient Database Explorer
Keep
Exclude
Export
6411/6411
14.4% opacity.
Hide Ticks
Dark
What is this?
A multidimensional explorer of nutrient data from the USDA.
The parallel coordinates displays the nutrient content of foods in the database across 14 dimensions, colored by food group.
Food Groups
327 Baby Foods
~~0 Baked Products~~
619 Beef Products
278 Beverages
Sample of 25 entries
Search Foods...
Alcoholic Beverage, wine, table, red, Gamay
Alcoholic beverage, distilled, whiskey, 86 proof
Bananas, raw

REMINDERS

3/2/2018 – Project 5 Due

3/7/2018 – Project 5 Peer Reviews Due

3/19/2018 – Paper Review Presentations Begin (look out for e-mail)

Hierarchical Parallel Coordinates for Exploration of Large Datasets

Ying-Huey Fua, Matthew O. Ward and Elke A. Rundensteiner
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
{yingfua,matt,rundenst}@cs.wpi.edu *

Abstract

Our ability to accumulate large, complex (multivariate) data sets has far exceeded our ability to effectively process them in search of patterns, anomalies, and other interesting features. Conventional multivariate visualization techniques generally do not scale well with respect to the size of the data set. The focus of this paper is on the interactive visualization of large multivariate data sets based on a number of novel extensions to the parallel coordinates display technique. We develop a multiresolutional view of the data via hierarchical clustering, and use a variation on parallel coordinates to convey aggregation information for the resulting clusters. Users can then navigate the resulting structure until the desired focus region and level of detail is reached, using our suite of navigational and filtering tools. We describe the design and implementation of our hierarchical parallel coordinates system which is based on extending the XmdvTool system. Lastly, we show examples of the tools and techniques applied to large (hundreds of thousands of records) multivariate data sets.

Keywords: Large-scale multivariate data visualization, hierarchical data exploration, parallel coordinates.

1 Introduction

- Dimensional embedding techniques, such as dimensional stacking [16] and worlds within worlds [6].
- Dimensional subsetting, such as scatterplots [5].
- Dimensional reduction techniques, such as multidimensional scaling [20, 15, 29], principal component analysis [12] and self-organizing maps [14].

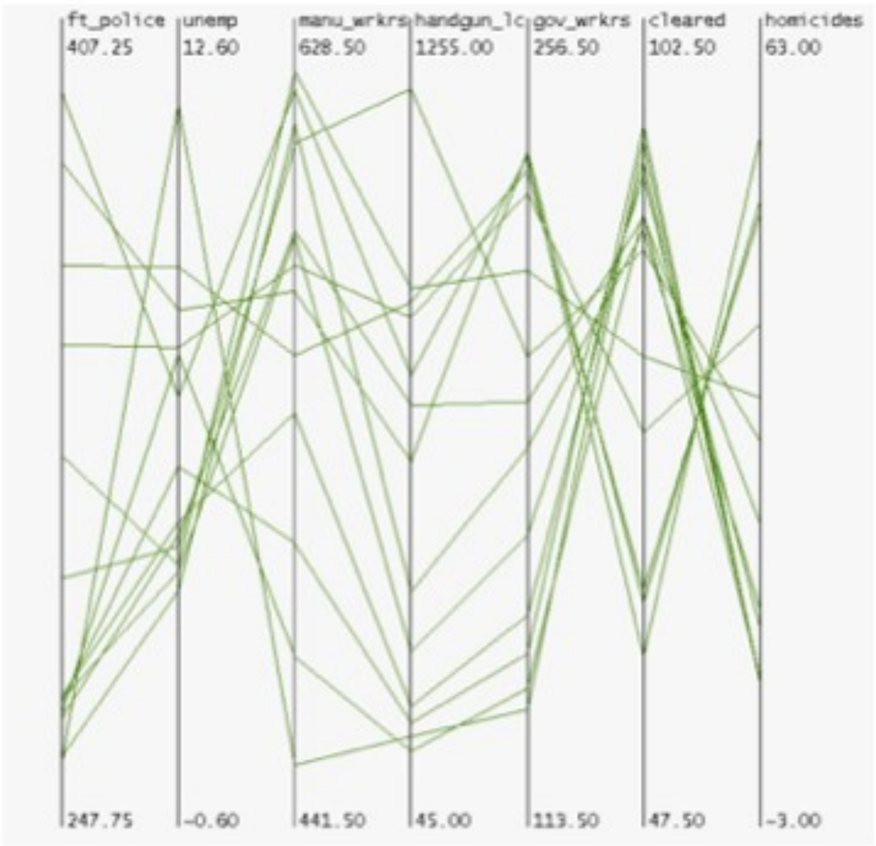
Most of these techniques do not scale well with respect to the size of the data set. As a generalization, we postulate that any method that displays a single entity per data point invariably results in overlapped elements and a convoluted display that is not suited for the visualization of large data sets. The quantification of the term "large" varies and is subject to revision in sync with the state of computing power. For our present application, we define a large data set to contain 10^6 to 10^9 data elements or more.

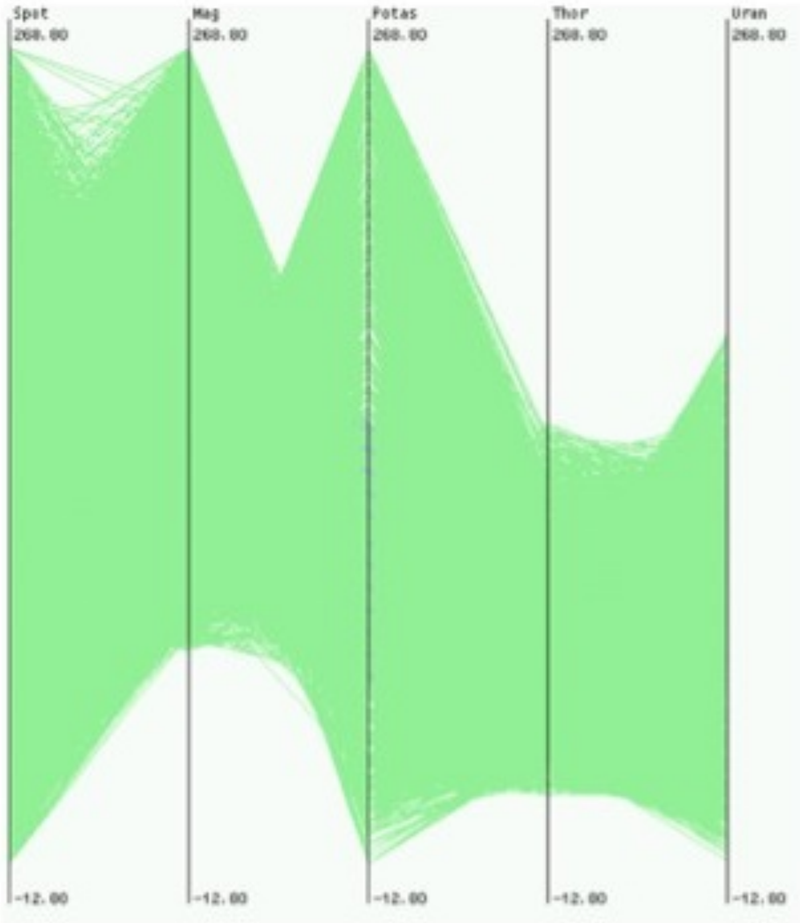
Our research focus extends beyond just data display, incorporating the process of data exploration, with the goal of interactively uncovering patterns or anomalies not immediately obvious or comprehensible. Our goal is thus to support an active process of discovery as opposed to passive display. We believe that it is only through data exploration that meaningful ideas, relations, and subsequent inferences may be extracted from the data. The major hurdles we need to overcome are the problems of display density/clutter (too

HIERARCHICAL PARALLEL COORDINATES

goal: scale up parallel coordinates to large datasets

challenge: overplotting/occlusion





FUA 1999

HPC: ENCODING DERIVED DATA

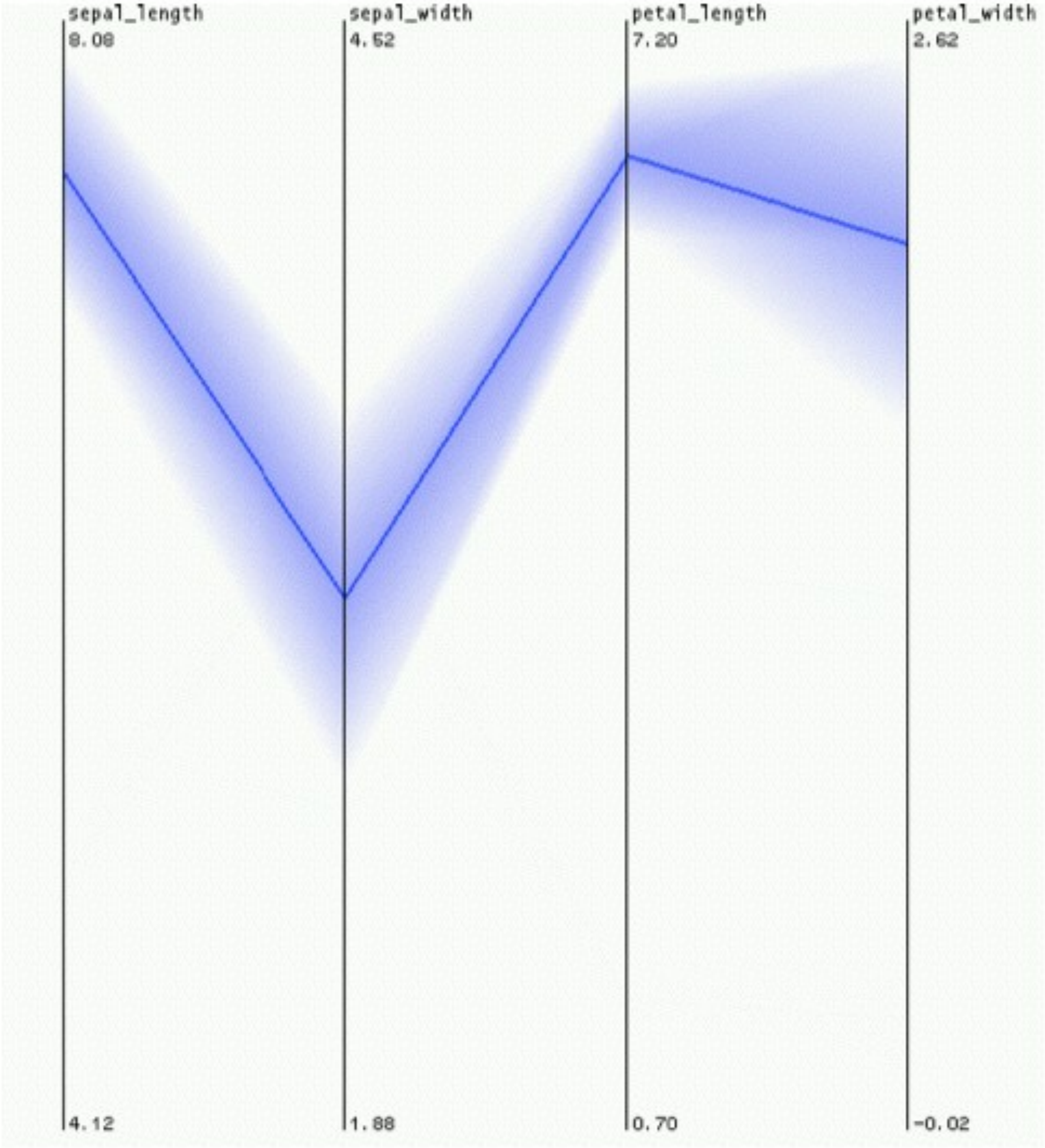
visual representation: variable-width opacity bands

show whole cluster, not just single item

min / max: spatial position

cluster density: transparency

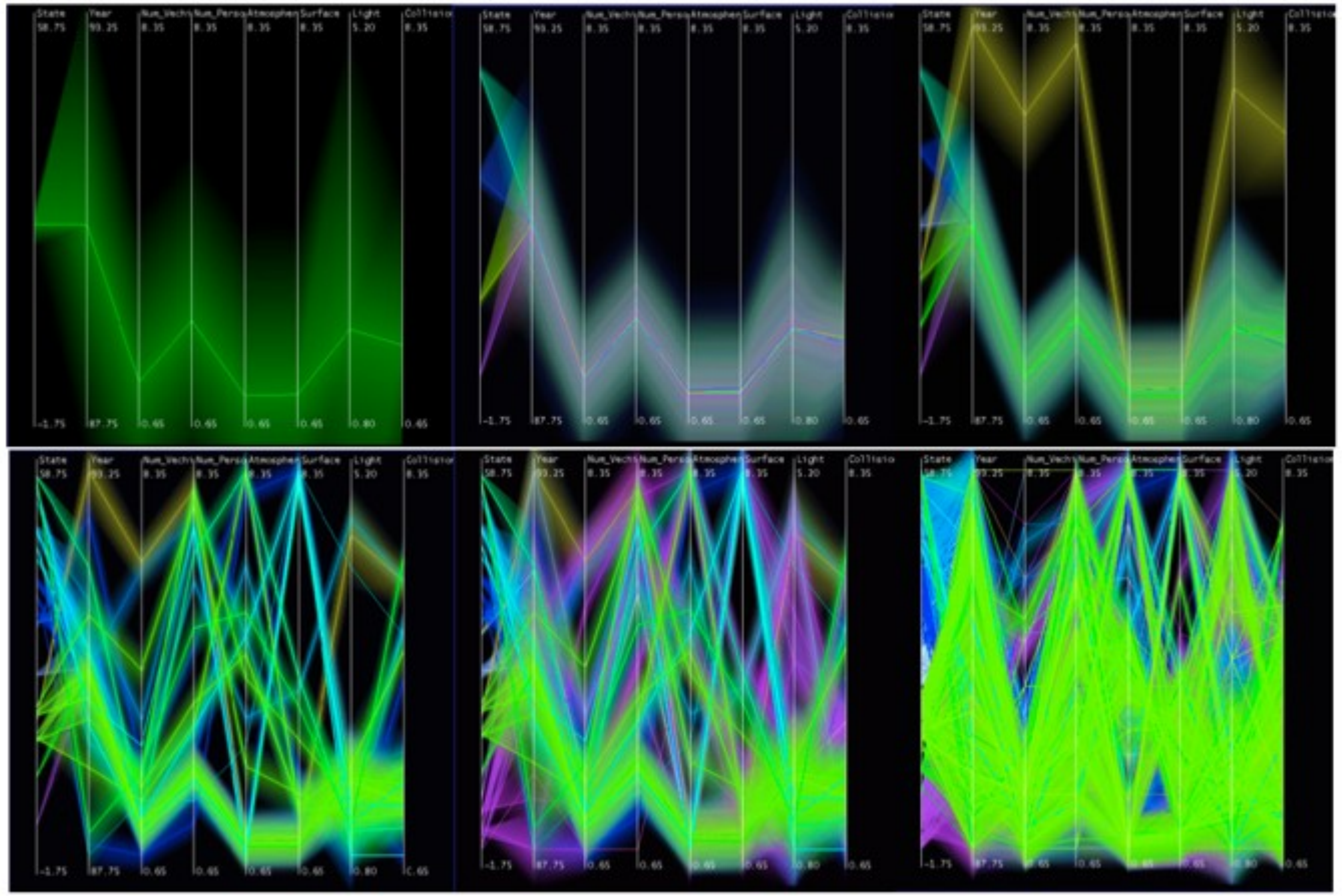
mean: opaque



FUA 1999

HPC: INTERACTING WITH DERIVED DATA

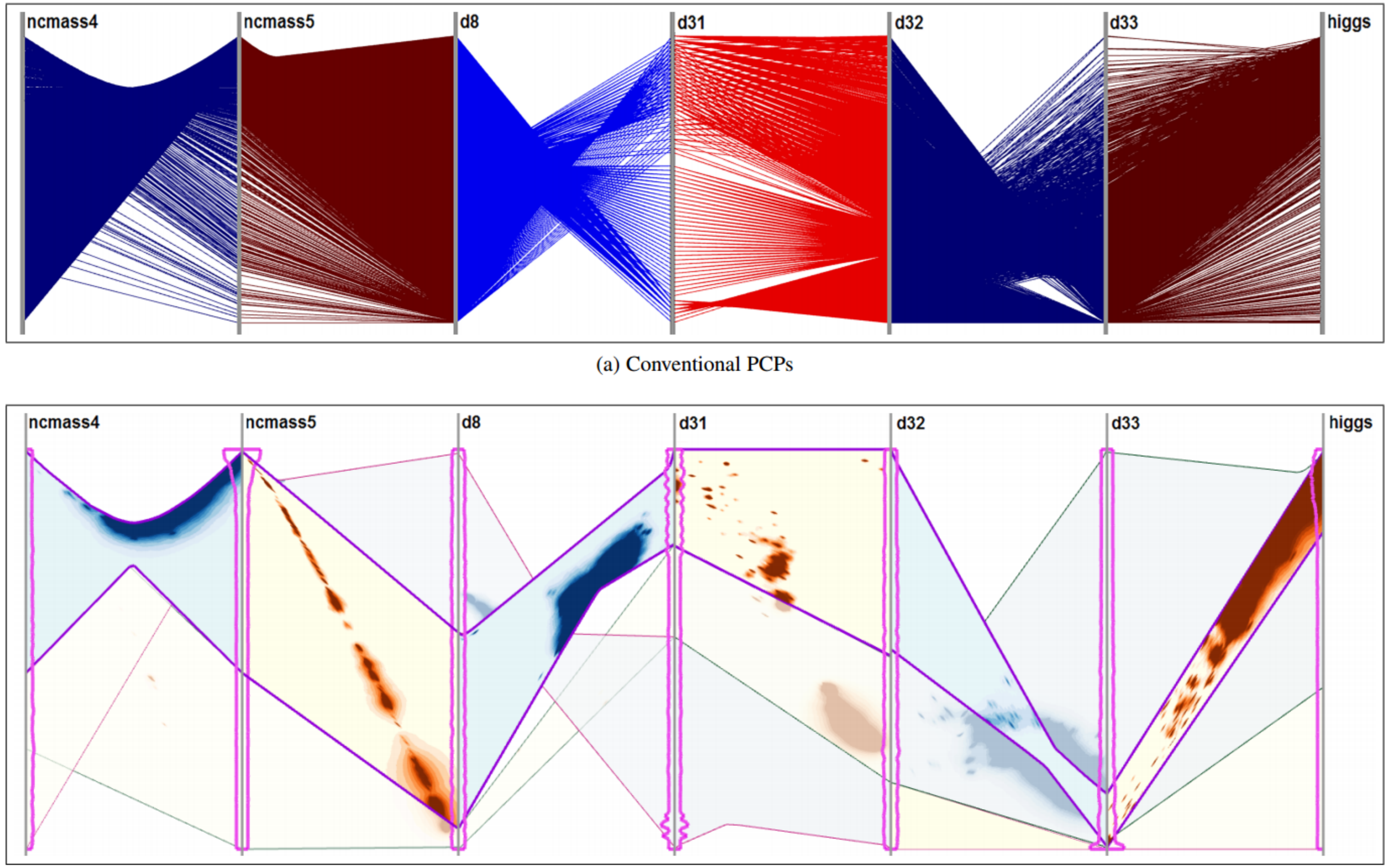
interactively change level of detail to navigate cluster hierarchy



FUA 1999

DSPCP

Cluster into groups of homogeneous behavior and represent positive and negative correlations directly



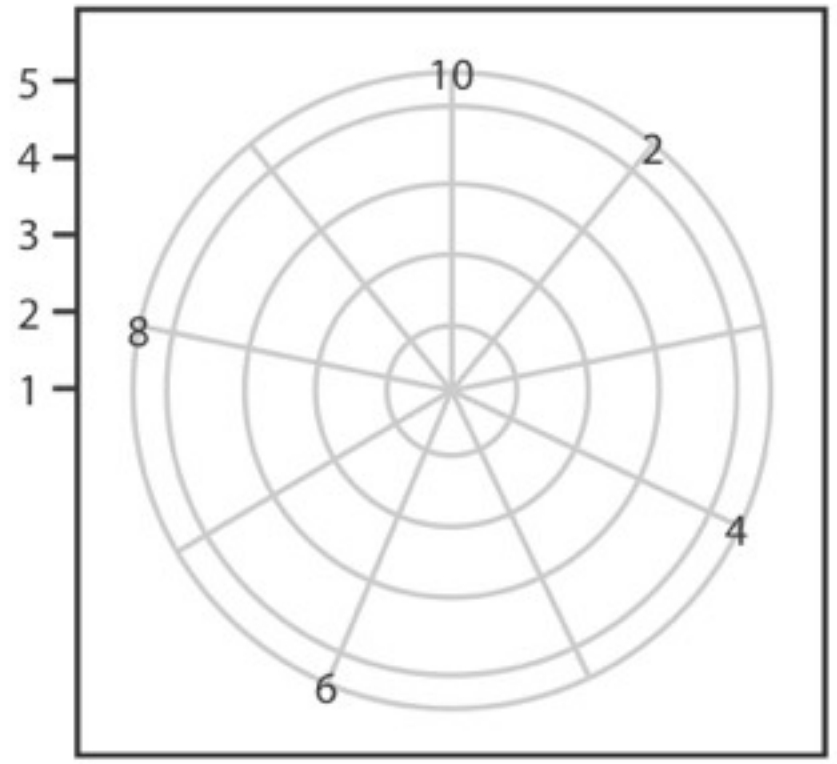
(a) Conventional PCPs

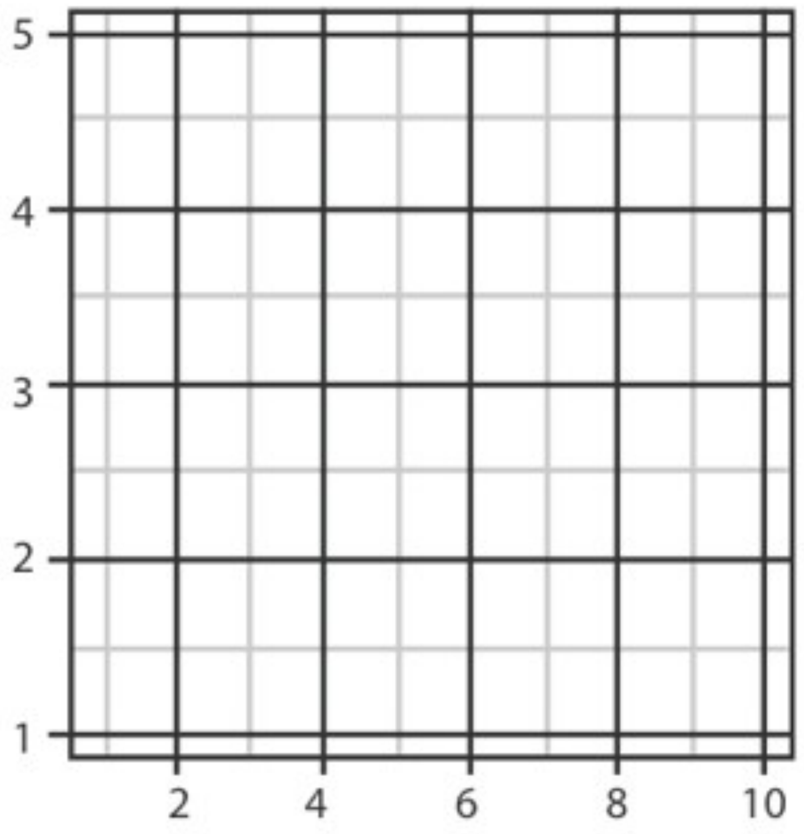
(b) DSPCP using K-means clustering

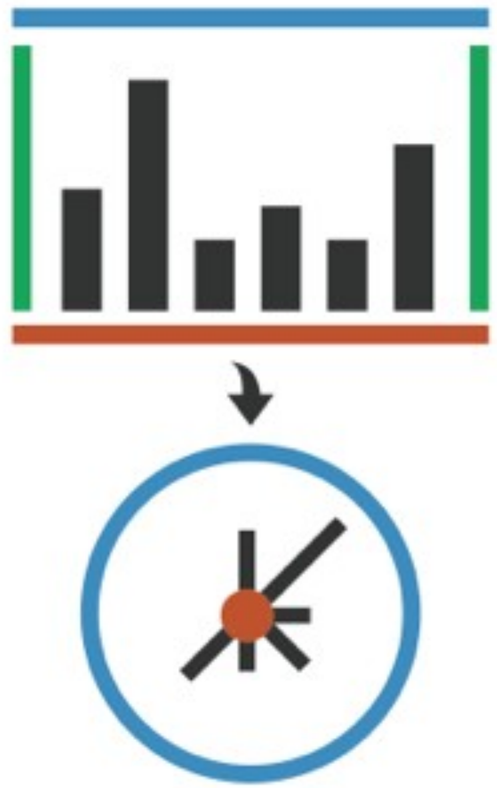
RADIAL LAYOUTS

HTTP://BL.OCKS.ORG/MBOSTOCK/3888852

RADIAL LAYOUTS USE POLAR COORDINATES



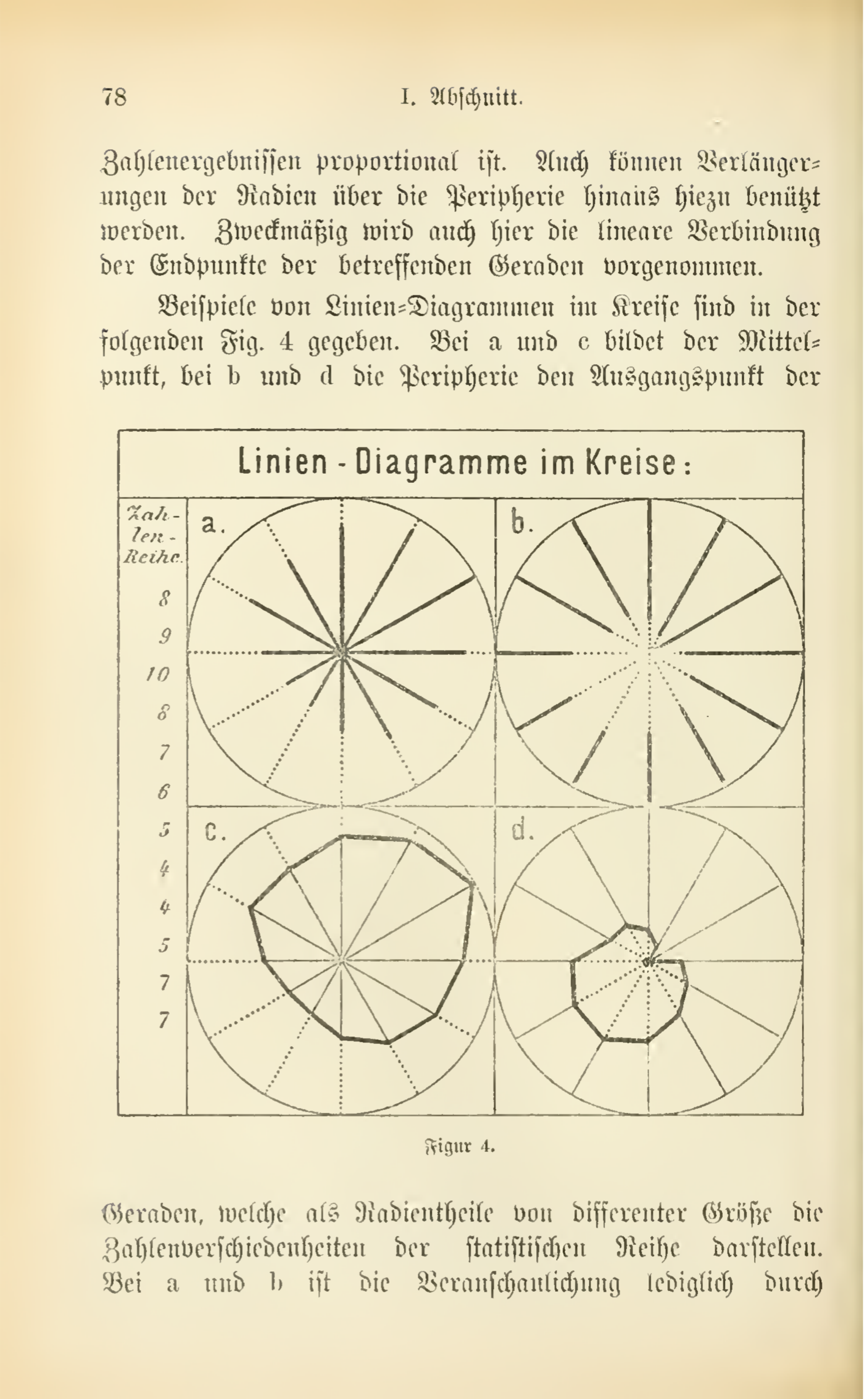




RADAR PLOT & STAR GRAPH

"parallel" dimensions in polar coordinate space

best if same units apply to each axis



78 I. Abschnitt.

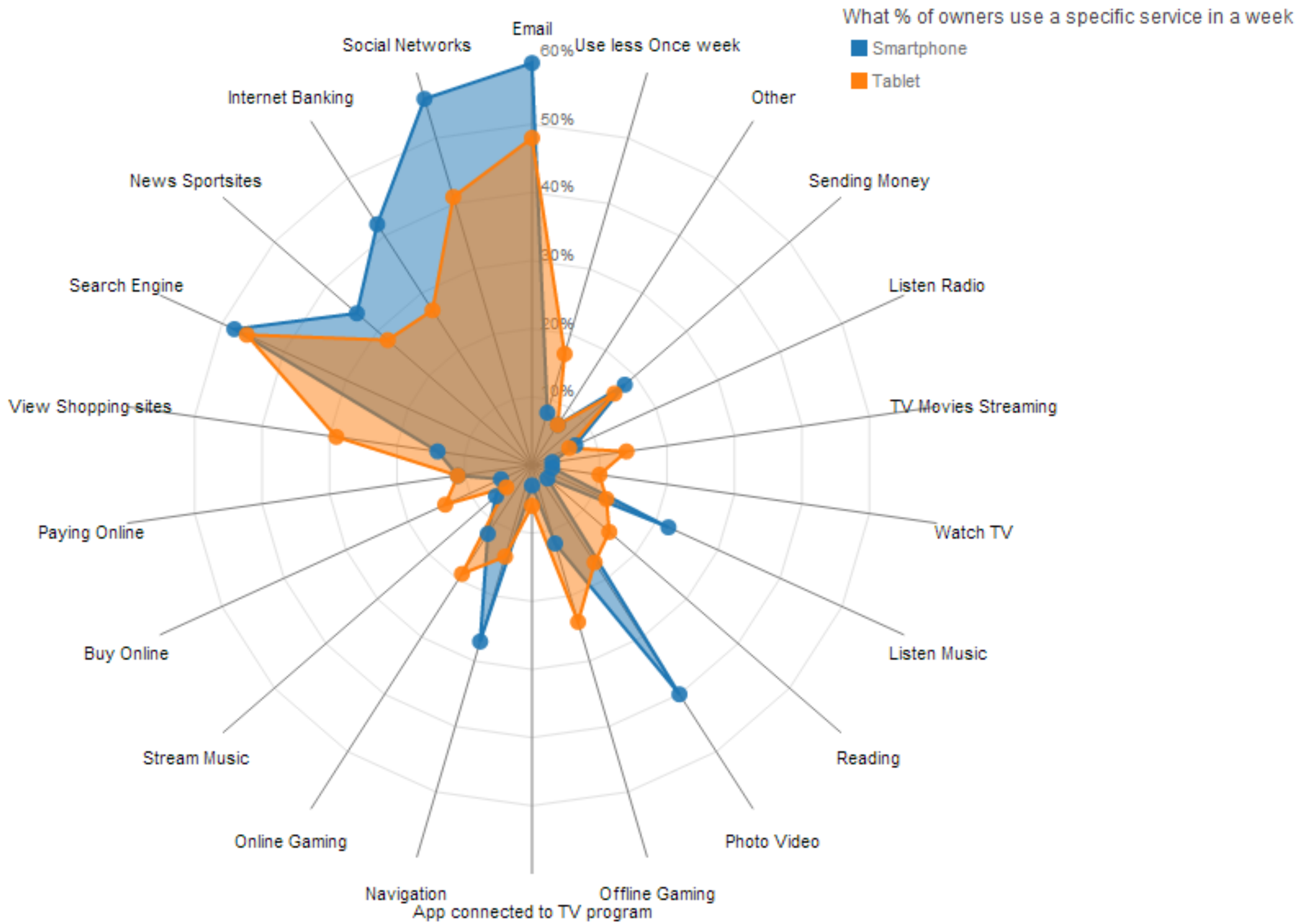
Zahlenergebnissen proportional ist. Auch können Verlängerungen der Radien über die Peripherie hinaus hiezu benützt werden. Zweckmäßig wird auch hier die lineare Verbindung der Endpunkte der betreffenden Geraden vorgenommen.

Beispiele von Linien-Diagrammen im Kreise sind in der folgenden Fig. 4 gegeben. Bei a und c bildet der Mittelpunkt, bei b und d die Peripherie den Ausgangspunkt der

Figur 4.

Geraden, welche als Radientheile von differenter Größe die Zahlenverschiedenheiten der statistischen Reihe darstellen. Bei a und b ist die Veranschaulichung lediglich durch

Critique: What do you think?



HTTP://WWW.VISUALCINNAMON.COM/2013/09/MAKING-D3-RADAR-CHART-LOOK-BIT-BETTER.HTML

DENSE PIXEL DISPLAY: VISDB

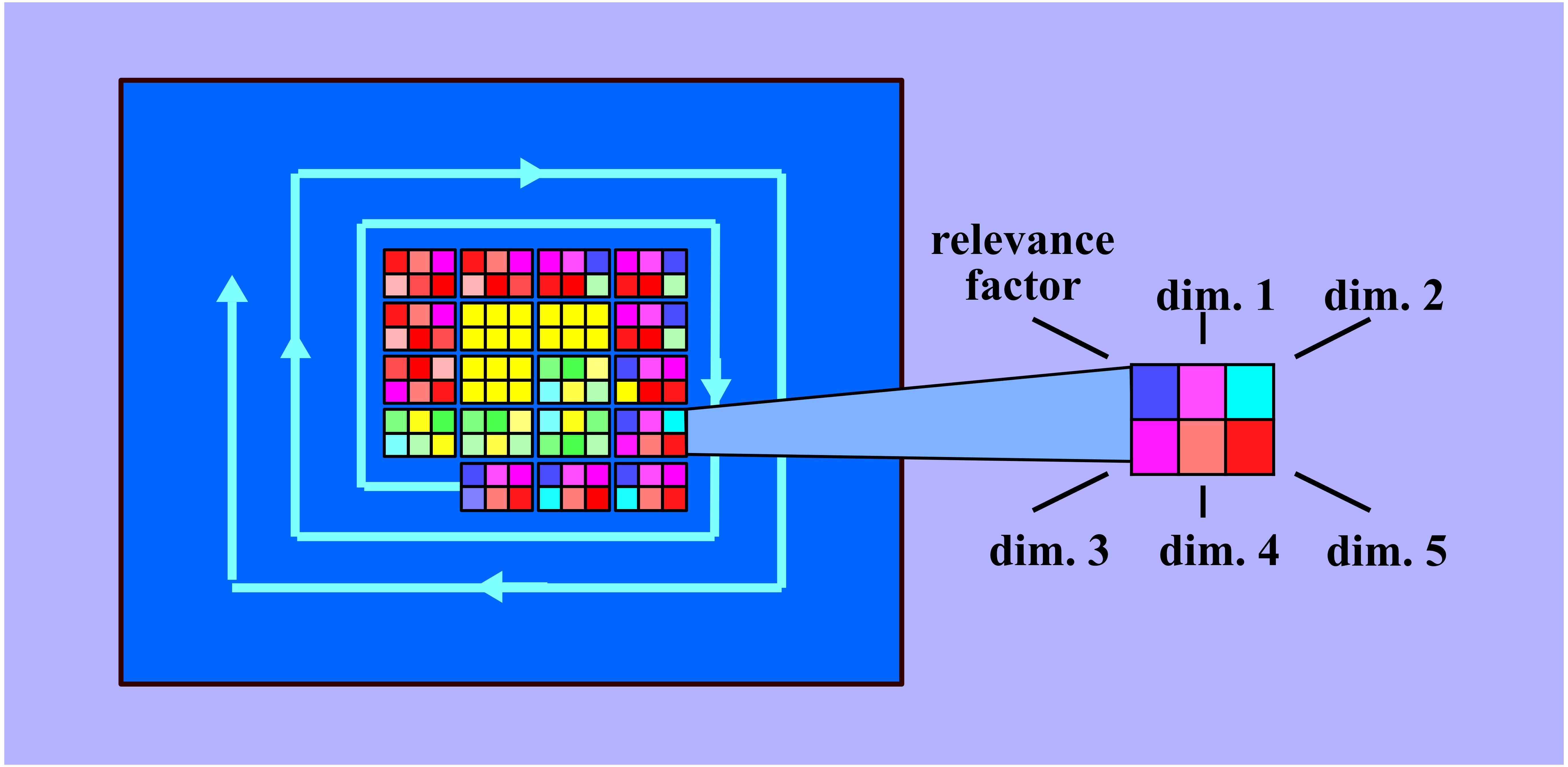
represent each data item, or each attribute in an item as a single pixel

can fit as many items on the screen as there are pixels, on the order of millions

relies heavily on color coding

challenge: what's the layout?

KEIM 1994


relevance factor
dim. 1
dim. 2
dim. 3
dim. 4
dim. 5

CRITIQUE: WHAT DO YOU THINK?