VISUALIZING MULTI-ATTRIBUTE DATA

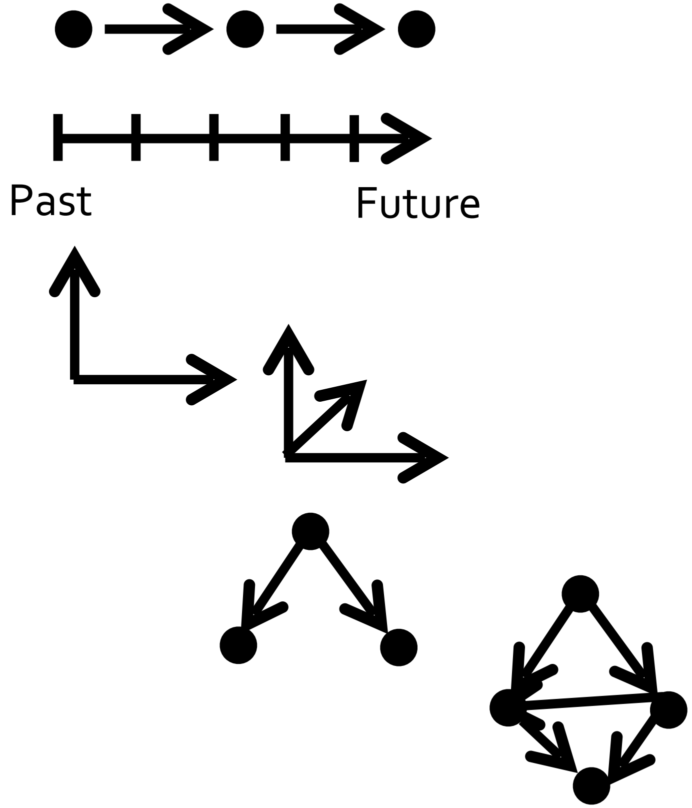
DATA TABLES

Petra Isenberg, Anastasia Bezerianos

informatics mathematics
Inria

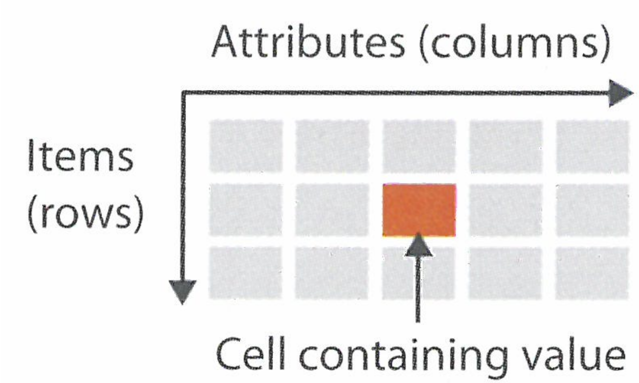
RECAP

- 1D (linear)
- Temporal
- 2D (maps)
- 3D
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)

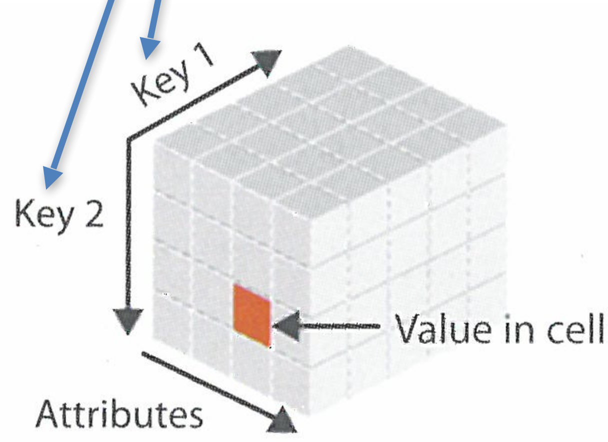


Shneiderman: The Eyes Have It

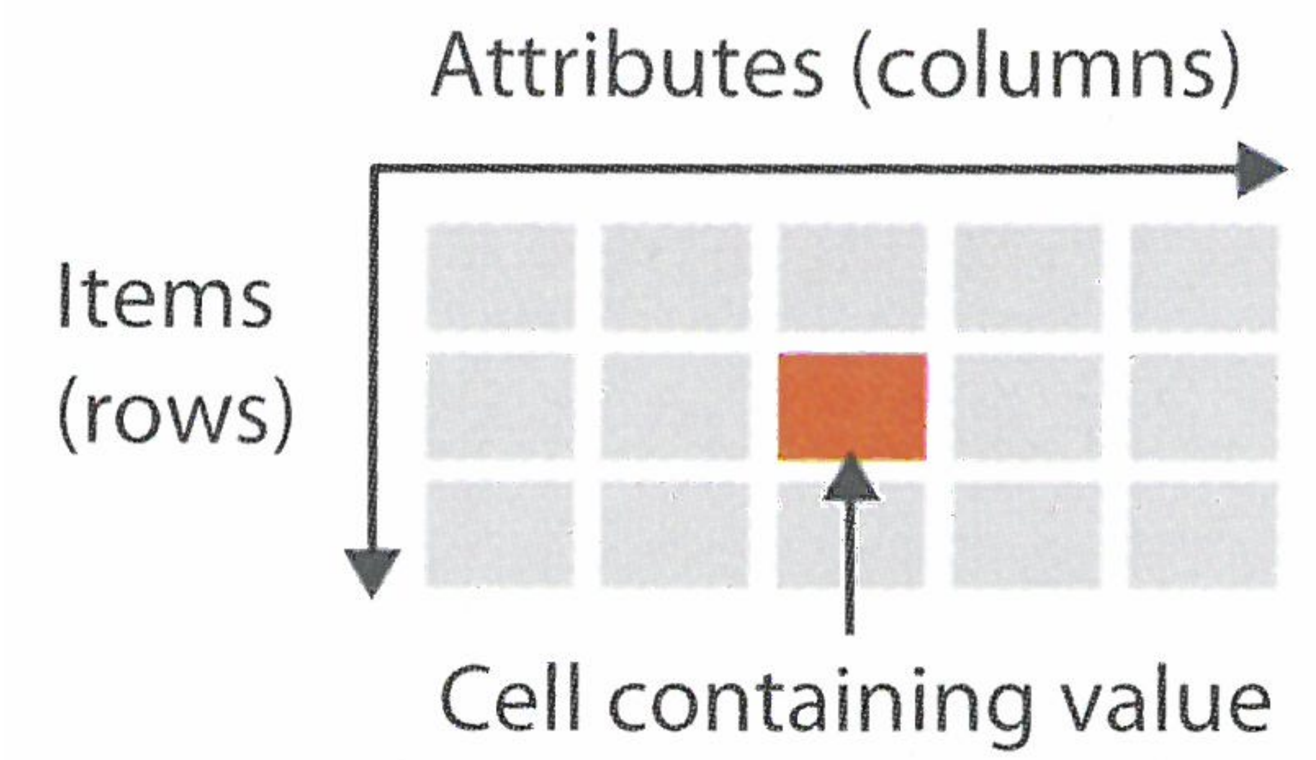
DATA TABLES - TERMINOLOGY



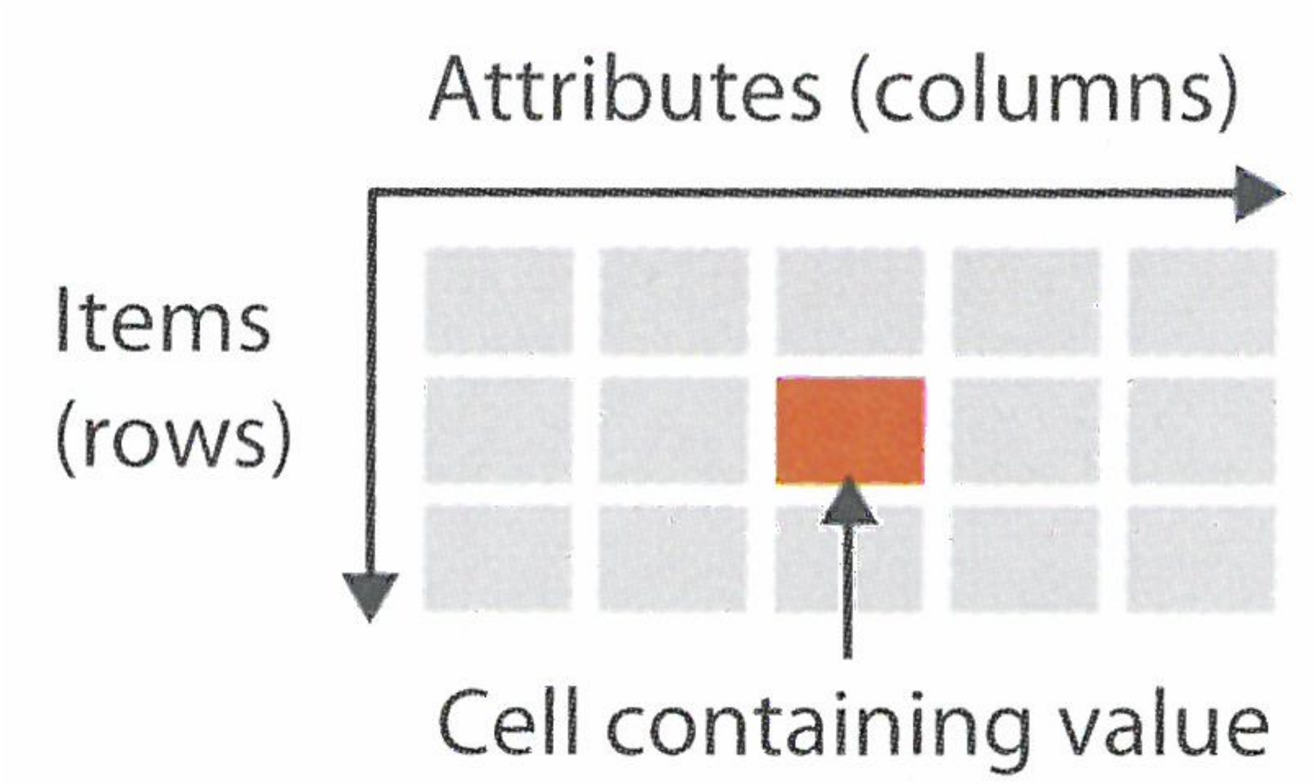
index to look up values



WHAT COULD BE THE KEY HERE?



WHAT DATA TYPE IS SUITABLE FOR A KEY?



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

Abc Vispubdata-Grobid-min-c... **Conference**	# Vispubdata-... **Year**	Abc Vispubdata-Grobid-min-clean.... **Paper.Title**
InfoVis	2015	A comparative study ...
InfoVis	2015	A Linguistic Approach...
InfoVis	2015	A Psychophysical Inv...
InfoVis	2015	A Simple Approach fo...
InfoVis	2015	Acquired Codes of Me...
InfoVis	2015	AggreSet: Rich and Sc...
InfoVis	2015	AmbiguityVis: Visuali...
InfoVis	2015	Automatic Selection ...
InfoVis	2015	Beyond Memorability...
InfoVis	2015	Beyond Weber's Law:...
InfoVis	2015	Evaluation of Parallel...
InfoVis	2015	Guidelines for Effecti...
InfoVis	2015	High-Quality Ultra-Co...
InfoVis	2015	HOLA: Human-like Ort...
InfoVis	2015	How do People Make ...

CONFERENCE:
InfoVis, Vis, SciVis, VAST

YEAR:
1990 – 2015

PAPER.TITLE:
>2500 different

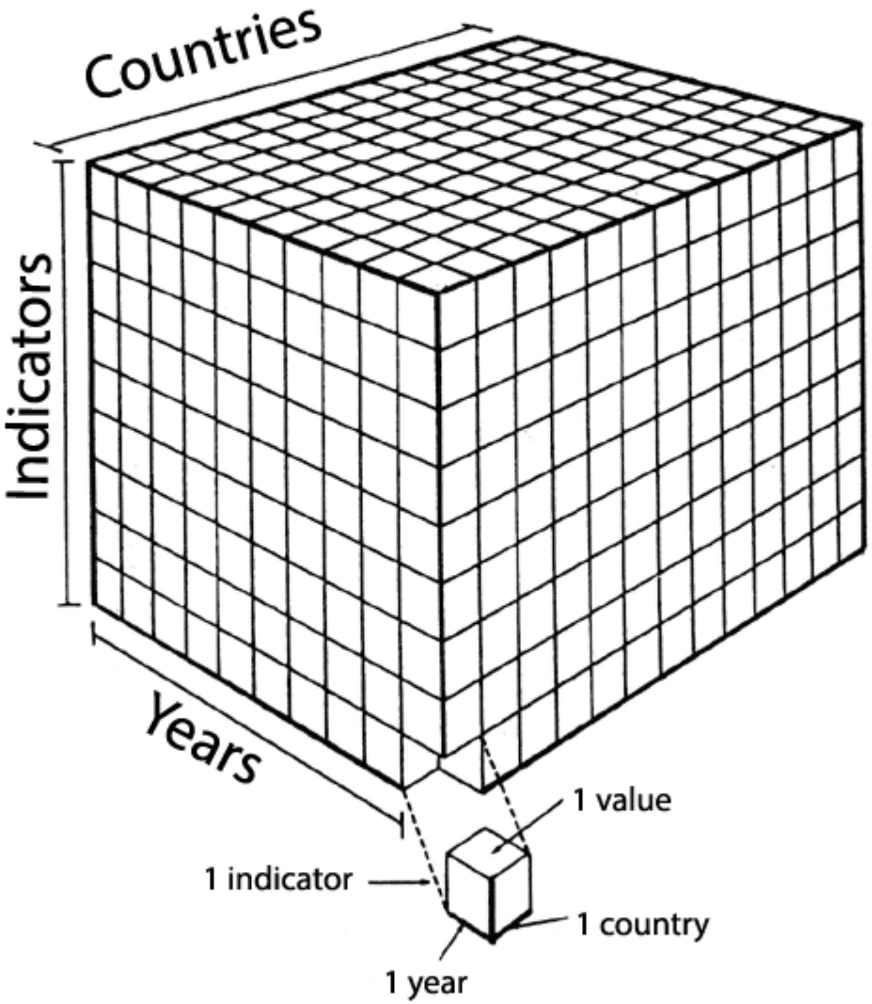
VISPUBDATA

ATTRIBUTES

ITEMS

C	Year	Paper.Title	Paper.DOI	Link	First.page	Last.page	Paper.type..C.conf...	Abstract	Author.Names	First.Author.Affilia...	Deduped.author.n...	References	Author.Keywords	OCR.Authors
In	2015	A comparative study ...	10.1109/TVCG.2015....	http://dx.doi.org/10....	619	628	J	RadViz and star coord...	Rubio-Sanchez, M.;Ra...	;;;	Rubio-Sanchez, M.;Ra...	10.1109/VAST.2010....	RadViz, Star coordina...	Rubio-S ´ Anchez,Ma...
In	2015	A Linguistic Approach...	10.1109/TVCG.2015....	http://dx.doi.org/10....	698	707	J	When data categorie...	Setlur, V.;Stone, M.C.	;	Setlur, V.;Stone, M.C.	null	linguistics, natural la...	Setlur,Vidya;Stone,M...
In	2015	A Psychophysical Inv...	10.1109/TVCG.2015....	http://dx.doi.org/10....	479	488	J	Physical visualization...	Jansen, Y.;Hornbaek, K.	Univ. of Copenhagen, ...	Jansen, Y.;Hornbaek, K.	10.1109/TVCG.2012....	Data physicalization, ...	Jansen,Yvonne;Hornb...
In	2015	A Simple Approach fo...	10.1109/TVCG.2015....	http://dx.doi.org/10....	678	687	J	General methods for ...	Simonetto, P.;Archam...	;;	Simonetto, P.;Archam...	10.1109/TVCG.2011....	Euler diagrams, Boun...	Simonetto,Paolo;Arc...
In	2015	Acquired Codes of Me...	10.1109/TVCG.2015....	http://dx.doi.org/10....	509	518	J	While information vis...	Byrne, L.;Angus, D.;W...	;;	Byrne, L.;Angus, D.;W...	10.1109/TVCG.2013....	Visual Design, Taxono...	Byrne,Lydia;Angus,D...
In	2015	AggreSet: Rich and Sc...	10.1109/TVCG.2015....	http://dx.doi.org/10....	688	697	J	Datasets commonly i...	Yalcin, M.A.;Elmqvist,...	Univ. of Maryland, Co...	Yalcin, M.A.;Elmqvist,...	10.1109/TVCG.2011....	Multi-valued attribut...	Adil Yalçın,M;Beders...
	2015	AmbiguityVis: Visuali...	10.1109/TVCG.2015....	http://dx.doi.org/10....	359	368	J	Node-link diagrams p...	Yong Wang;Qiaomu S...	;;;;;;	Yong Wang;Qiaomu S...	10.1109/TVCG.2006....	Visual Ambiguity, Vis...	Wang,Yong;Shen,Qia...
InfoVi	2015	Automatic Selection ...	10.1109/TVCG.2015....	http://dx.doi.org/10....	669	677	J	Effective small multi...	Anand, A.;Talbot, J.	;	Anand, A.;Talbot, J.	10.1109/VAST.2010....	Small multiple displa...	Anand,Anushka;Talbo...
InfoVis	2015	Beyond Memorability...	10.1109/TVCG.2015....	http://dx.doi.org/10....	519	528	J	In this paper we mov...	Borkin, M.A.;Bylinskii...	;;;;;;;	Borkin, M.;Bylinskii, Z...	10.1109/TVCG.2012....	Information visualiza...	null
InfoVis	2015	Beyond Weber's Law:...	10.1109/TVCG.2015....	http://dx.doi.org/10....	469	478	J	Models of human per...	Kay, M.;Heer, J.	;	Kay, M.;Heer, J.	10.1109/TVCG.2014....	Weber's law, percept...	Kay,Matthew;Heer,Je...
InfoVis	2015	Evaluation of Parallel...	10.1109/TVCG.2015....	http://dx.doi.org/10....	579	588	J	The parallel coordina...	Johansson, J.;Forsell,...	Norrkoping Visualiza...	Johansson, J.;Forsell,...	10.1109/TVCG.2014....	Survey, evaluation, g...	Johansson,Jimmy;For...
InfoVis	2015	Guidelines for Effecti...	10.1109/TVCG.2015....	http://dx.doi.org/10....	489	498	J	Semi-automatic text ...	Strobelt, H.;Oelke, D.;...	;;;;	Strobelt, H.;Oelke, D.;...	10.1109/TVCG.2012....	Text highlighting tec...	Strobelt,Hendrik;Oel...
InfoVis	2015	High-Quality Ultra-Co...	10.1109/TVCG.2015....	http://dx.doi.org/10....	339	348	J	Prior research into ne...	Yoghourdjian, V.;Dwy...	;;;;;	Yoghourdjian, V.;Dwy...	10.1109/TVCG.2008....	Network visualizatio...	Yoghourdjian,Vahan;...
InfoVis	2015	HOLA: Human-like Ort...	10.1109/TVCG.2015....	http://dx.doi.org/10....	349	358	J	Over the last 50 year...	Kieffer, S.;Dwyer, T.;...	;;;	Kieffer, S.;Dwyer, T.;...	10.1109/TVCG.2006....	Graph layout, orthog...	Kieffer,Steve;Dwyer,...
InfoVis	2015	How do People Make ...	10.1109/TVCG.2015....	http://dx.doi.org/10....	499	508	J	In this paper, we wou...	Sukwon Lee;Sung-He...	Sch. of Ind. Eng., Purd...	Sukwon Lee;Sung-He...	10.1109/TVCG.2013....	Sensemaking model, i...	Lee,Sukwon;Kim,Sun...
InfoVis	2015	Improving Bayesian R...	10.1109/TVCG.2015....	http://dx.doi.org/10....	529	538	J	Decades of research ...	Ottley, A.;Peck, E.M.;...	;;;;;;;	Ottley, A.;Peck, E.M.;...	10.1109/TVCG.2014....	Bayesian Reasoning, ...	Ottley,Alvitta;Peck,E...
InfoVis	2015	Matches, Mismatche...	10.1109/TVCG.2015....	http://dx.doi.org/10....	449	458	J	The energy performa...	Brehmer, M.;Ng, J.;Ta...	;;;	Brehmer, M.;Ng, J.;Ta...	10.1109/TVCG.2011....	Design study, design ...	Brehmer,Matthew;N...

THE DATA CUBE



Country	Year	Child mortality	Births per woman
Afghanistan	2014	68.1	4.8
Afghanistan	2013	69.9	5.1
France	2014	3.6	2.0
France	2013	3.6	2.0
USA	2014	5.7	5.9
USA	2013	1.9	1.9

MULTI-ATTRIBUTE DATA – OUR VIEW TODAY

n x d matrix

n attributes

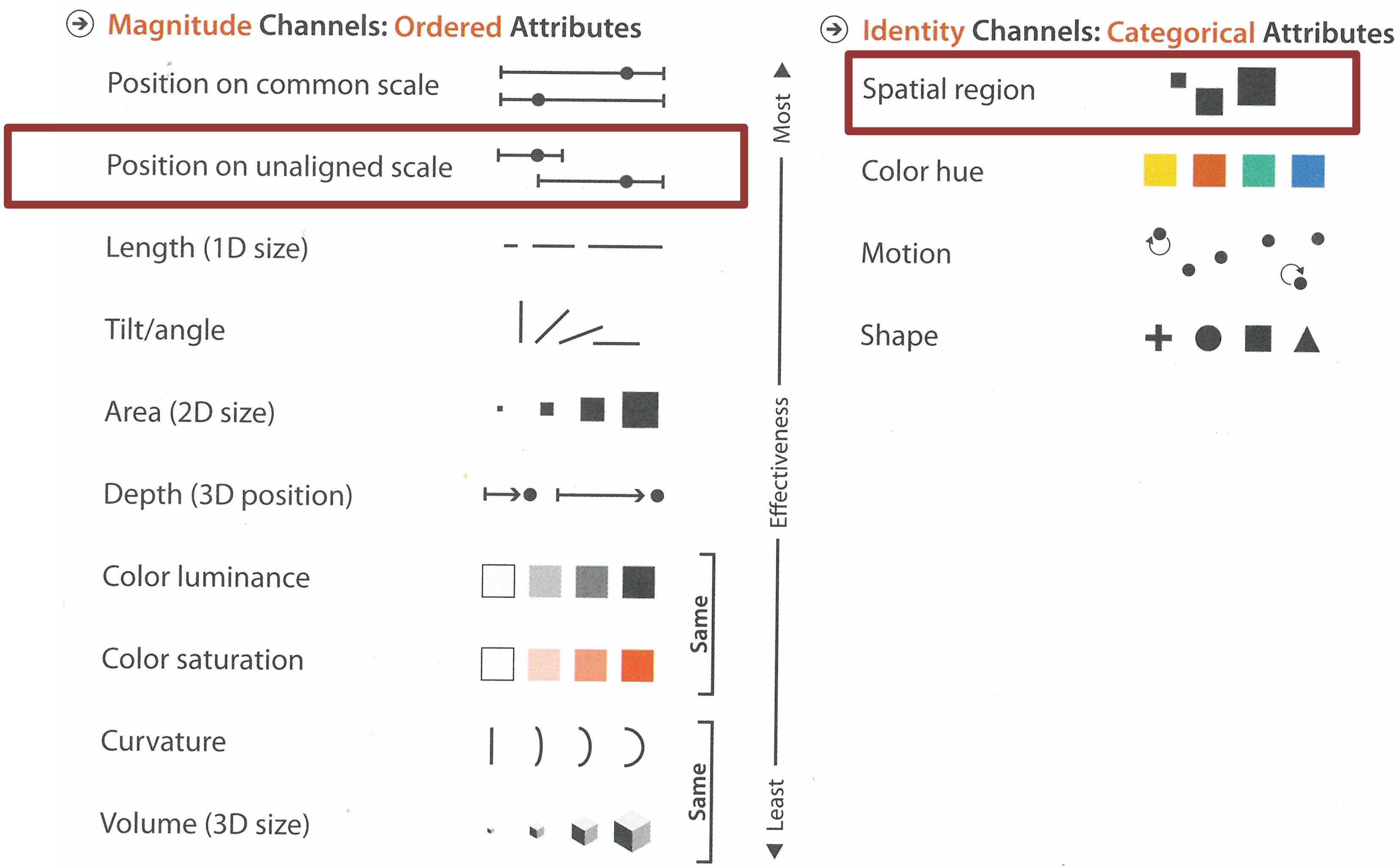
d items (data points)

Country	Year	Child mortality	Births per woman
Afghanistan	2014	68.1	4.8
Afghanistan	2013	69.9	5.1
France	2014	3.6	2.0
France	2013	3.6	2.0
USA	2014	5.7	5.9
USA	2013	1.9	1.9

ARRANGING TABULAR DATA

In Space

ARRANGING DATA



QUANTITATIVE VALUES

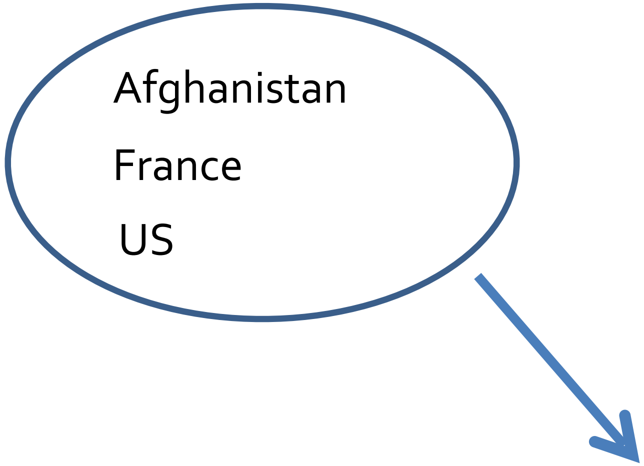
APPROACH

- Let's start with two attributes: country & income per person

Country	Income per person
Afghanistan	850
France	29500
US	41000

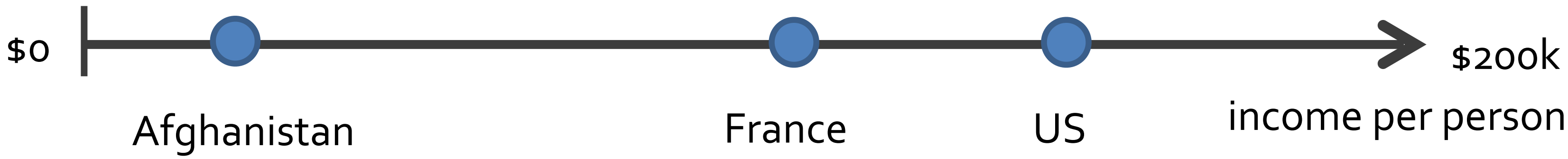
1. FIND A LAYOUT

Country	Income per person
Afghanistan	850
France	29500
US	41000



2. CHOOSE A VISUAL ENCODING & MARK

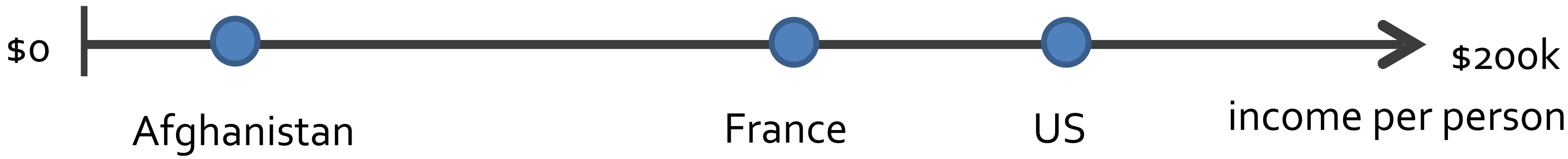
E.g. position + circle



1. FIND A LAYOUT

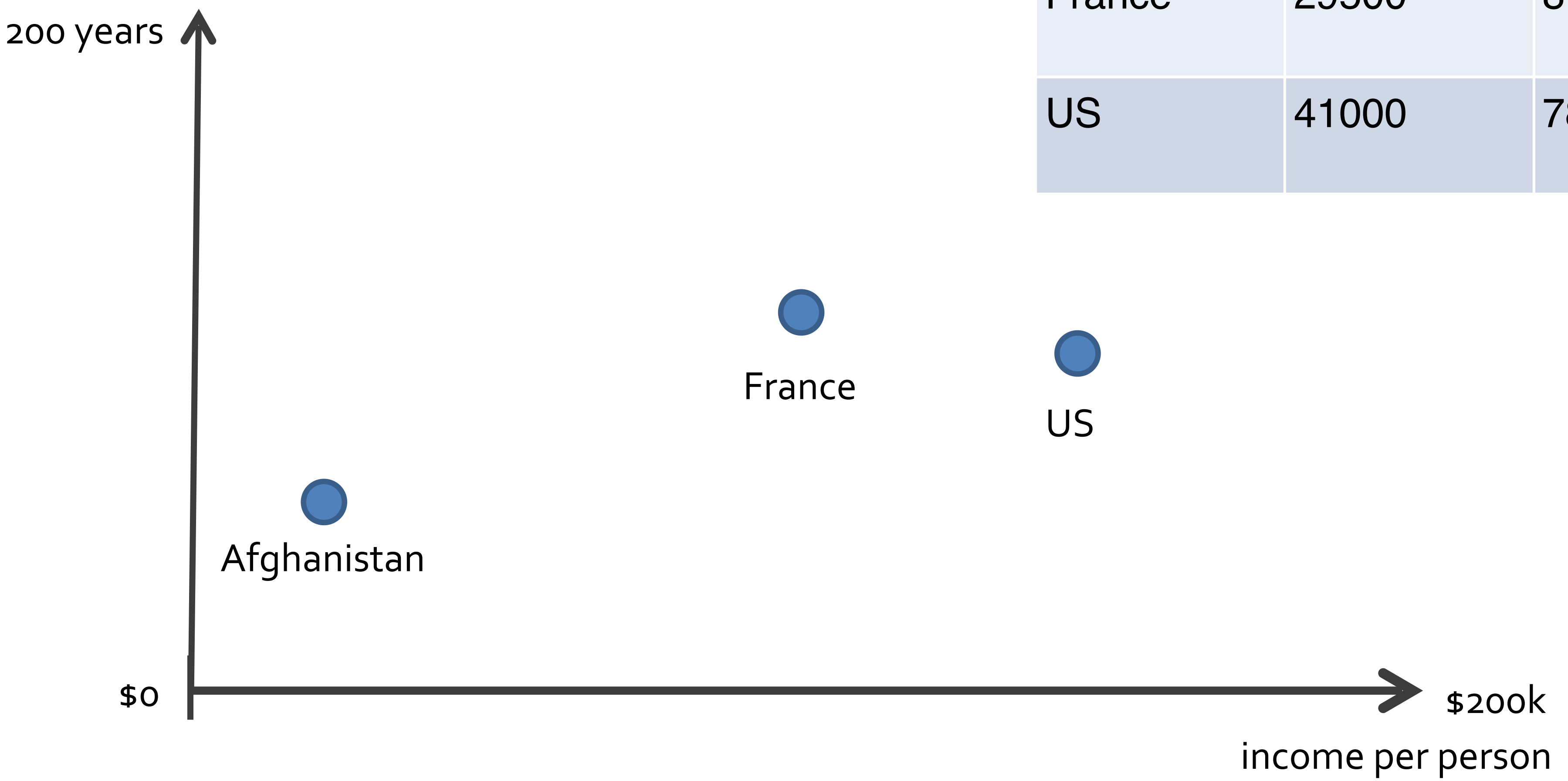
Country	Income per person	Life expectancy
Afghanistan	850	57
France	29500	81
US	41000	78

How do we extend this to 3 data attributes?



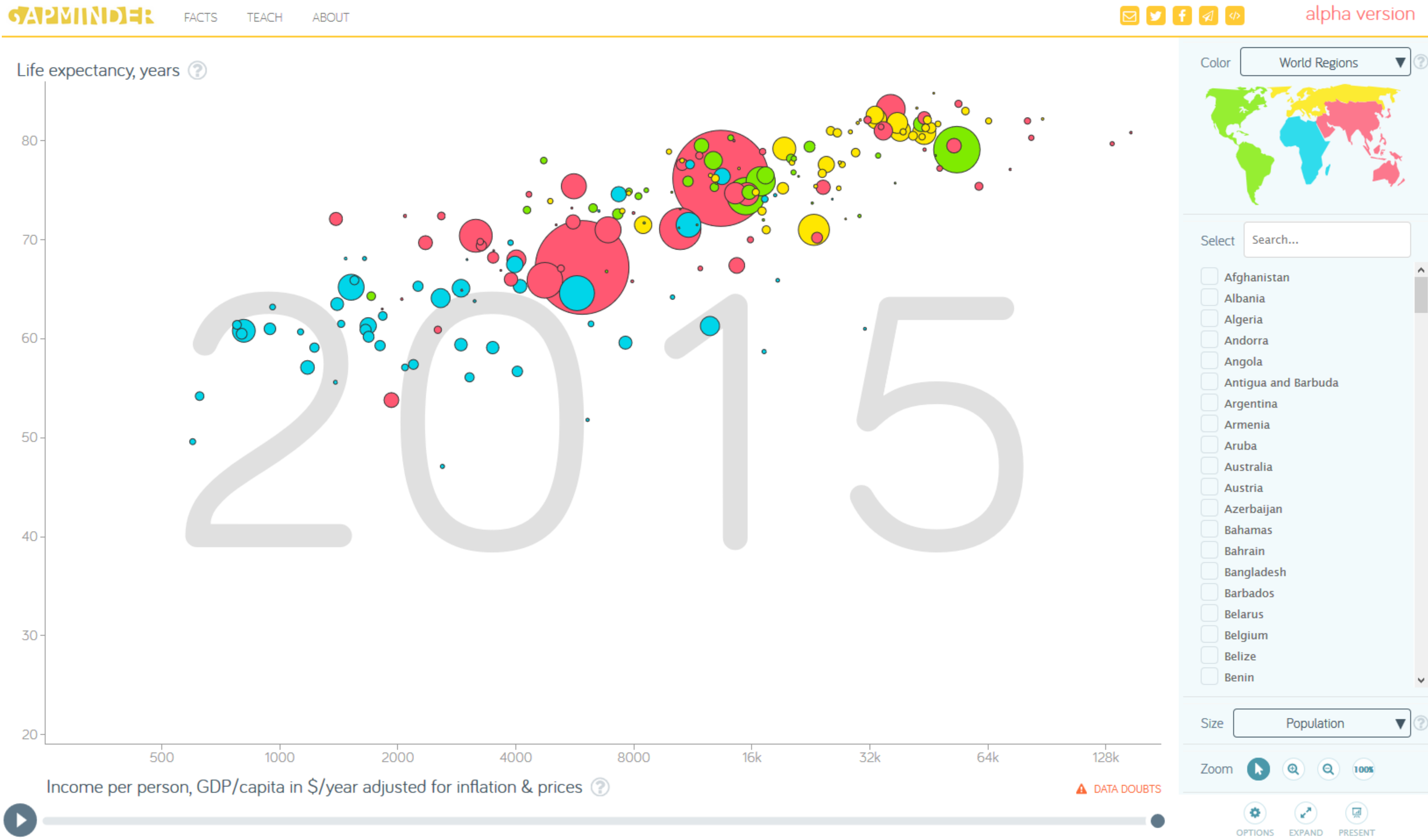
1. FIND A LAYOUT

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



SCATTERPLOTS

- two quantitative values
- horizontal and vertical spatial dimensions
- mark type = point

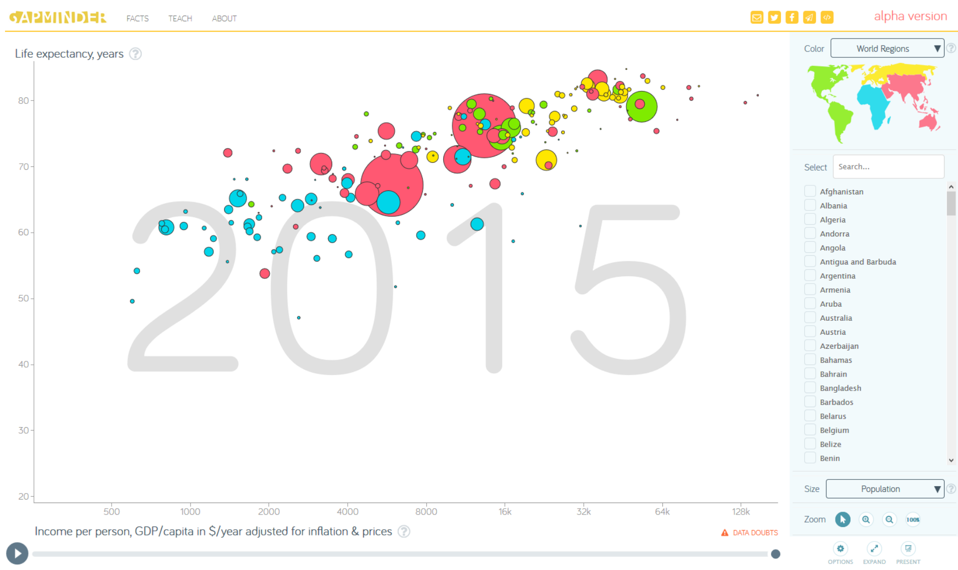


when marks are sized, the chart is often called a bubble chart or bubble plot

https://www.gapminder.org/

TASKS

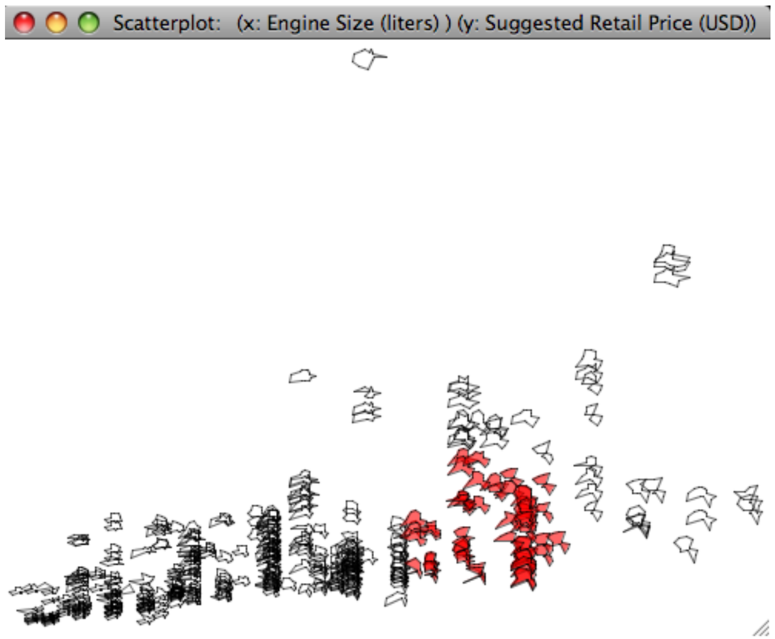
- find trends
- find outliers
- show distribution
- show correlation
- locate clusters



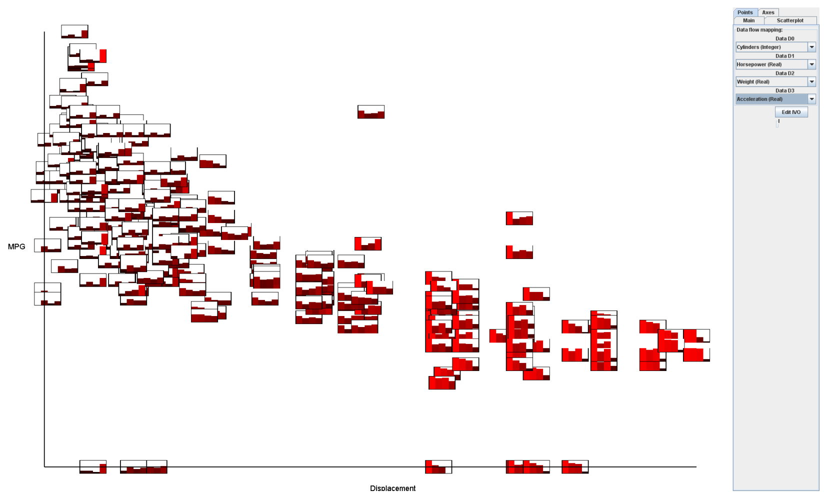
GLYPHS

marks can be replaced with glyphs

glyphs are themselves composed of multiple marks



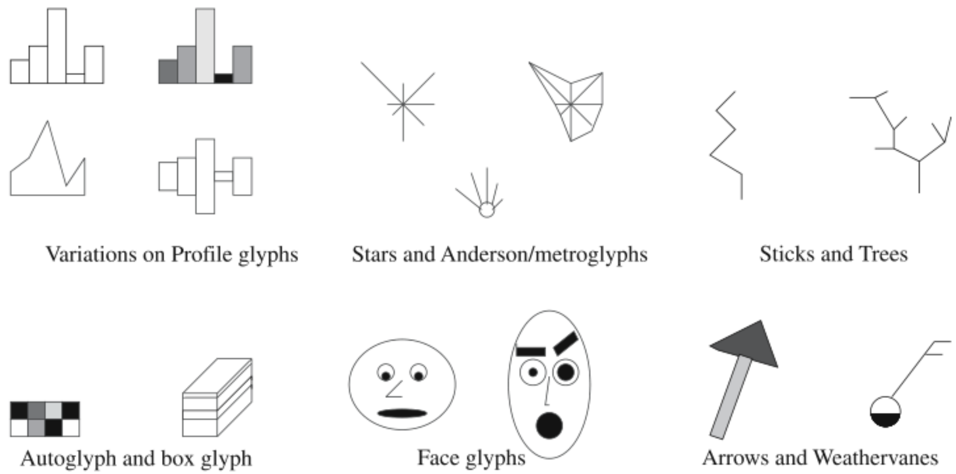
http://rosuda.org/software/Gauguin/gauguin.html



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

GLYPHS

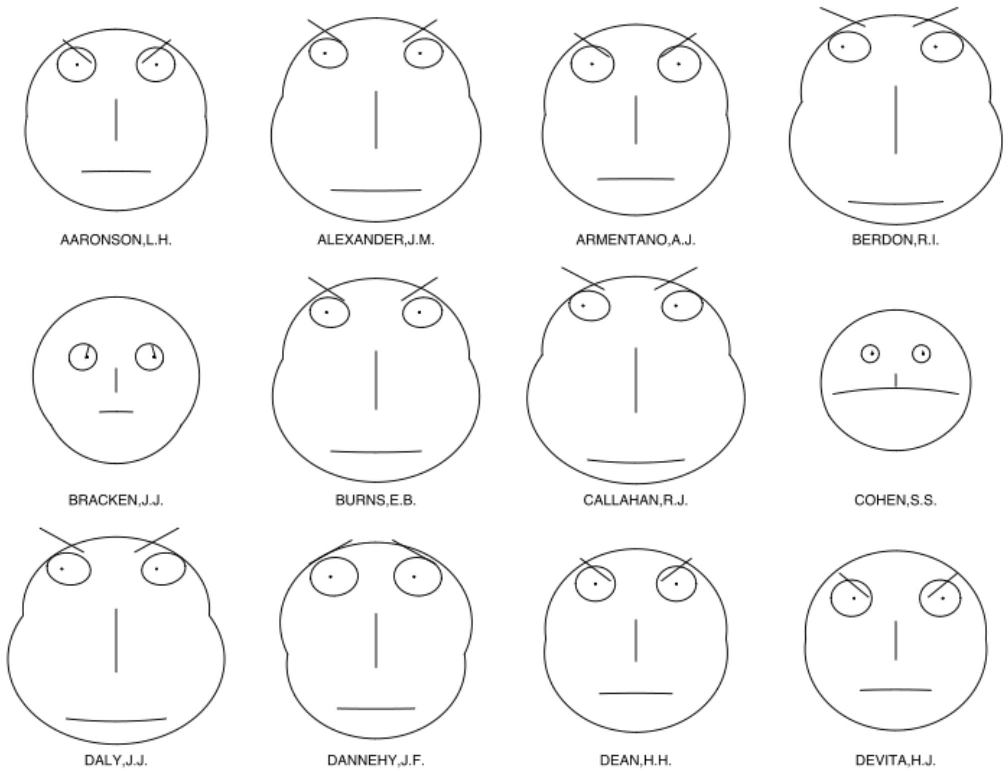
- Small composite visual representations of multi-dimensional data points
- Characterized generally by lack of reference structures (grid lines, axes labels, …)



From Ward, 2002
A taxonomy of glyph placement strategies for multidimensional data visualization

CHERNOFF FACES

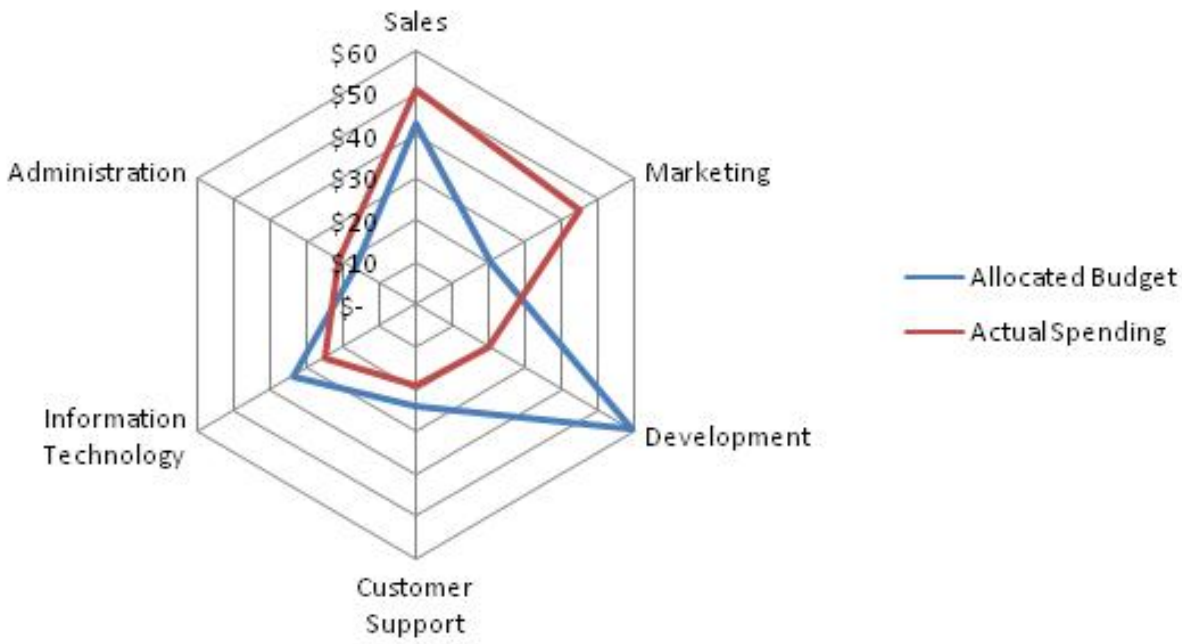
- features of a human face encode data values (e.g. slant of eye brows, size of eyes, …)
- reasoning: humans are good at differentiating faces and reading face features
- problem: chernoff faces have generally been found not to be very effective



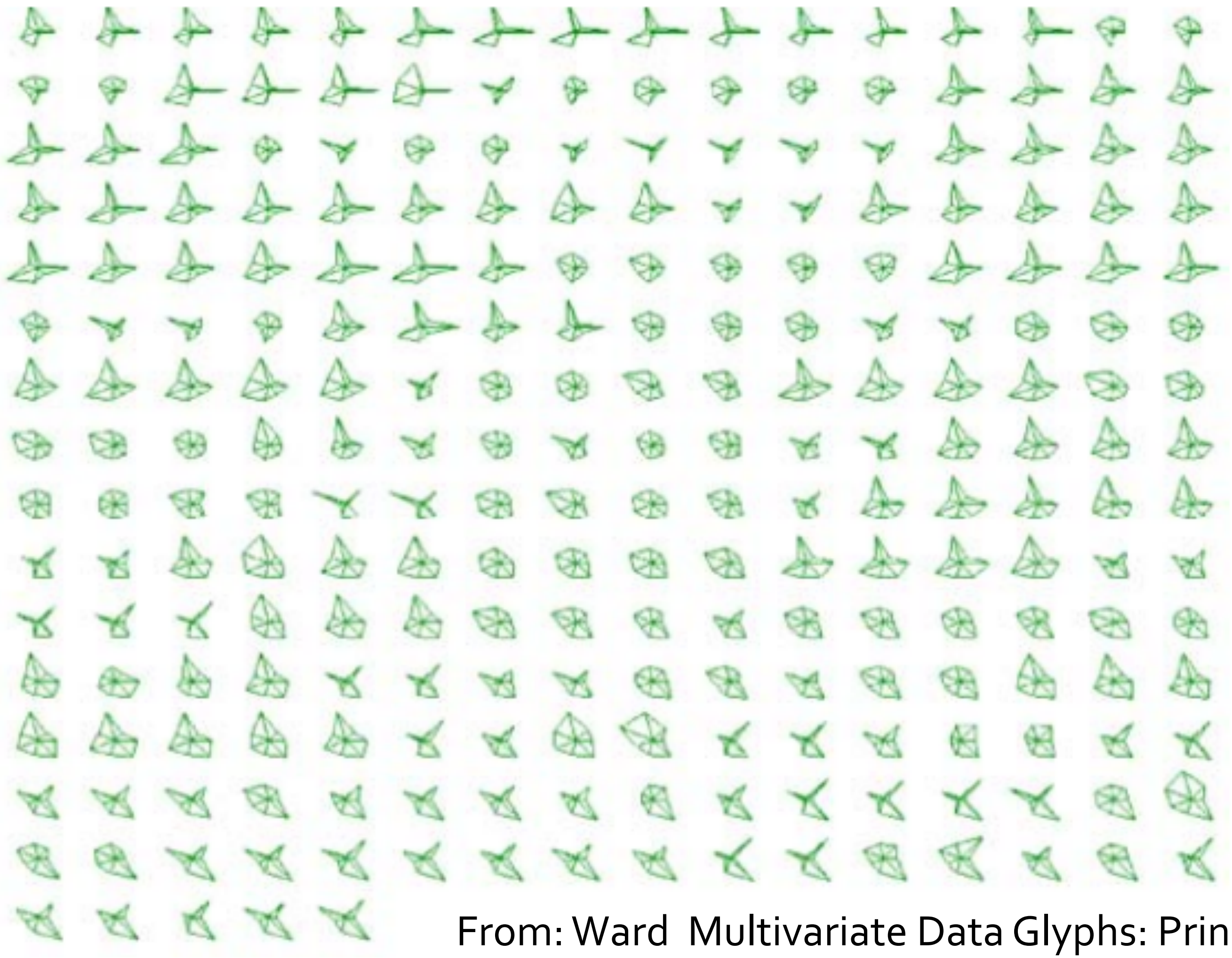
Herman Chernoff, [The Use of Faces to Represent Points in K-Dimensional Space Graphically](). 1973.

EXAMPLE: STAR GLYPHS

- Lay out dimension in radial fashion
- Draw each point as a ring



STAR GLYPHS



From: Ward Multivariate Data Glyphs: Principles and Practice. Handbook of Data Visualization (2008)

SHOW CATEGORICAL REGIONS

Separate, Order, and Align

CATEGORICAL VALUES

- 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

REGIONS

- contiguous bounded areas
- distinct from one another
- need to be separated, ordered, and aligned

LIST ALIGNMENT

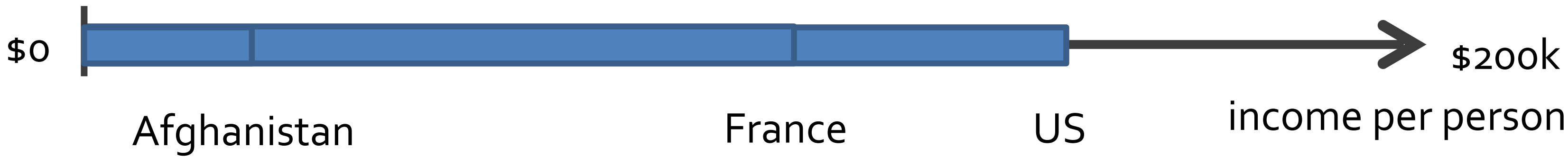
ONE KEY

LIST ALIGNMENT

separate into regions by key

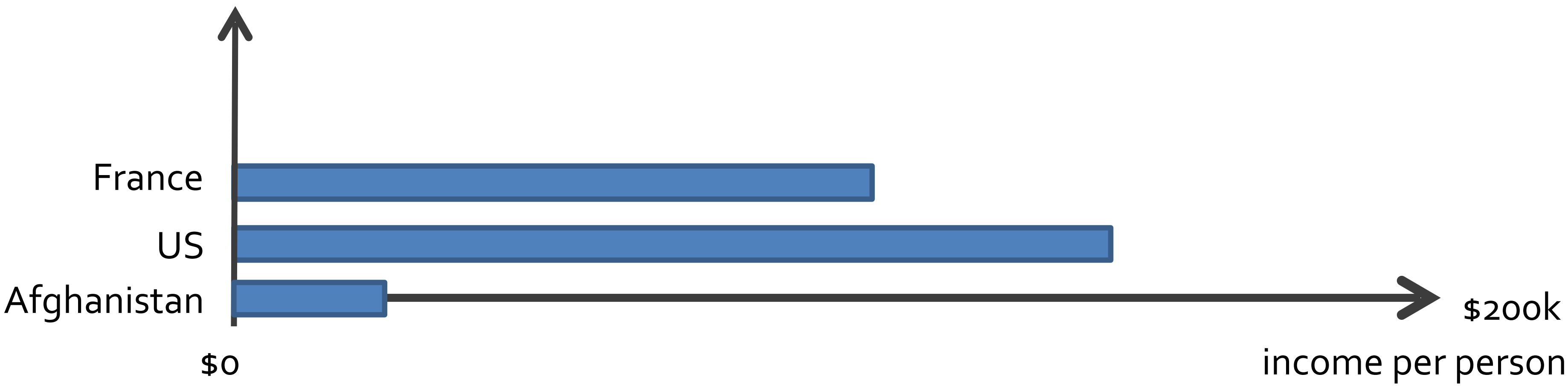
E.g. length + rectangle

Ok, this is weird



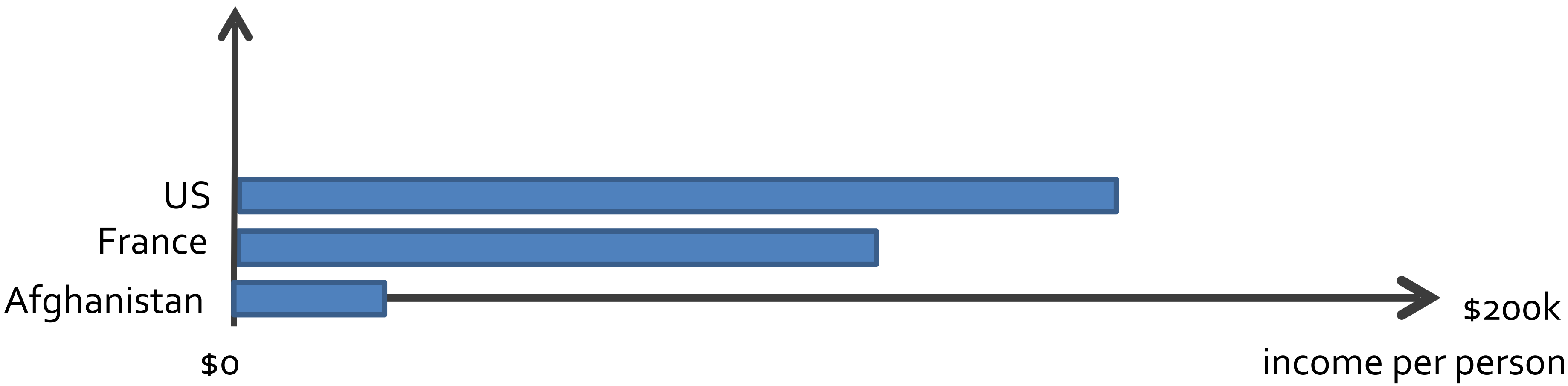
ALIGN

align regions of key categorical values along one axis in a common frame



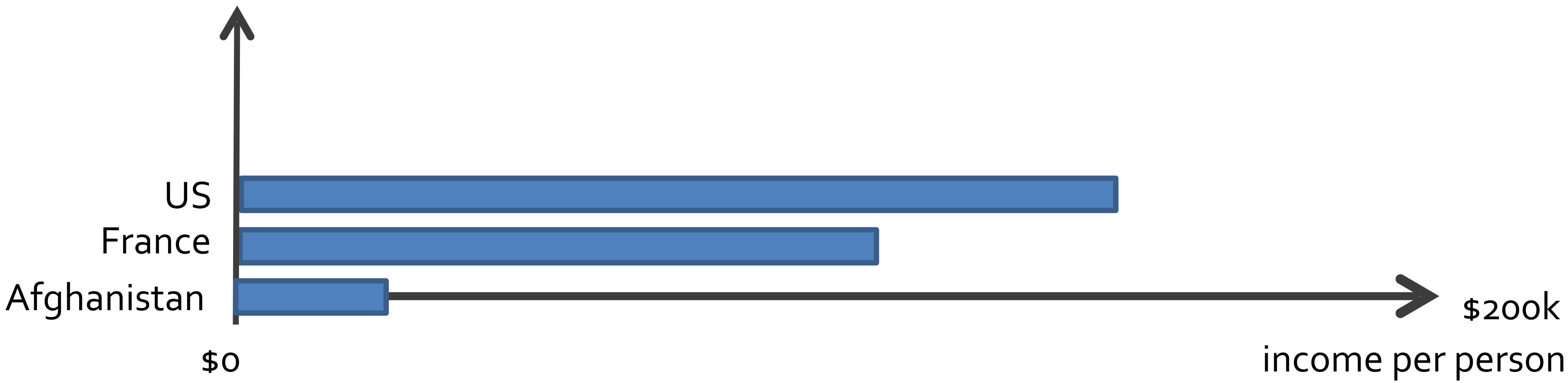
ORDER

- using a derived attribute such as alphabet
- and/or using dependent data values

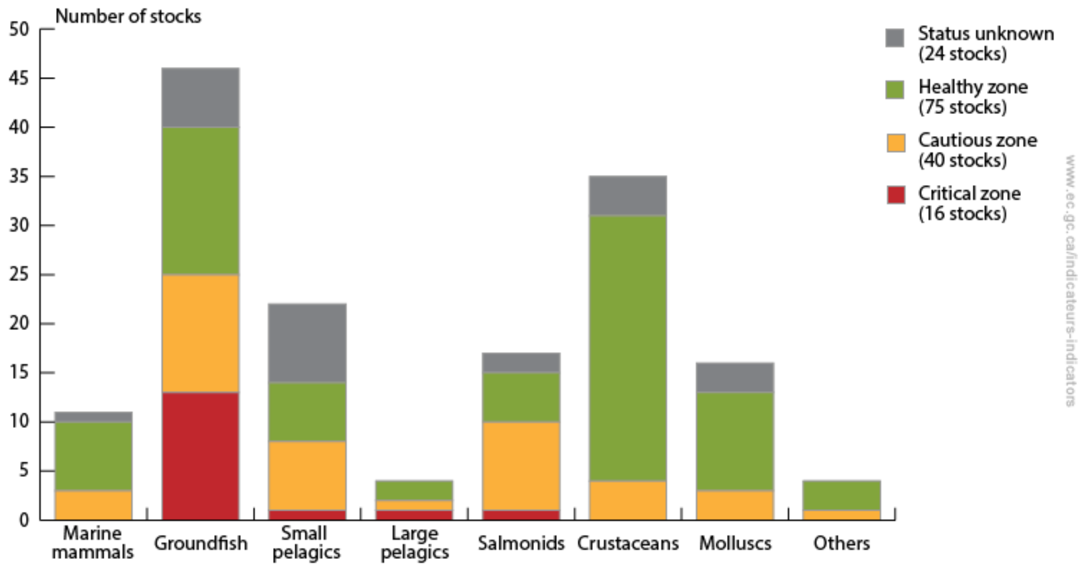


BAR CHARTS

DATA	one quantitative value attribute, one categorical key attribute
ENCODE	line marks, express value attribute with aligned vertical position (length), separate key attribute with horizontal position
TASK	lookup and compare values
SCALE	key attribute: dozens to hundreds of levels



ALTERNATIVE ALIGNMENT

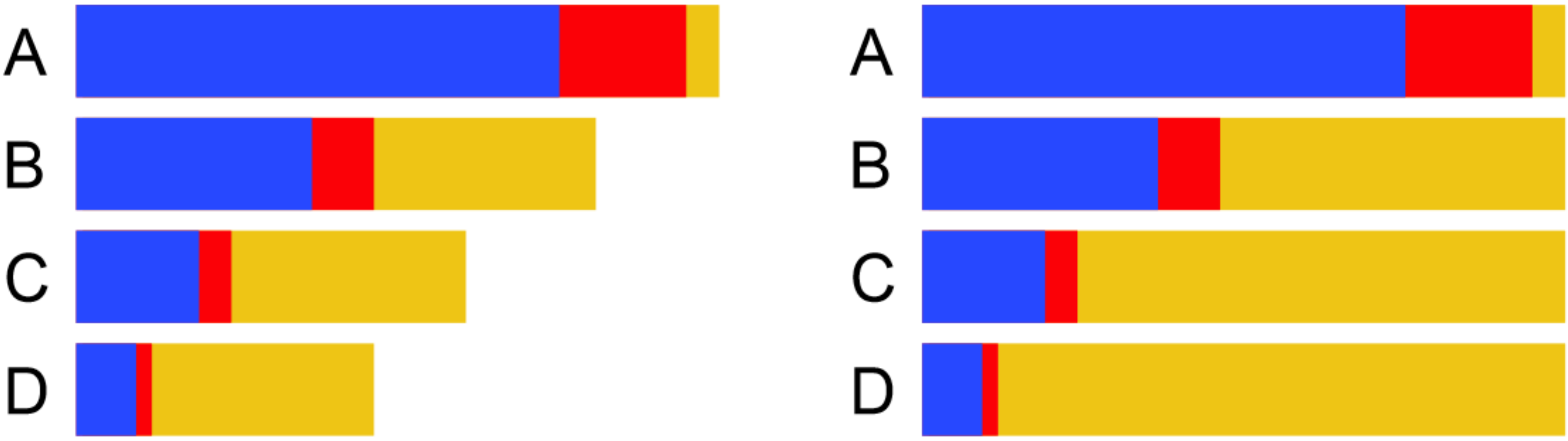


https://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=1BCD421B-1

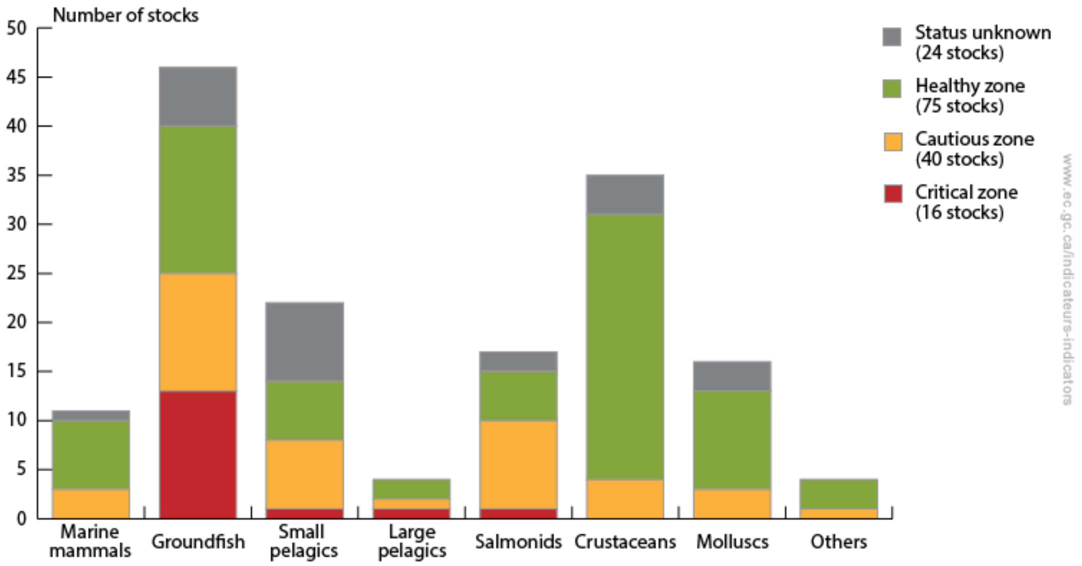
Stacked bar chart

- 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

STACKED BARS VS. NORMALIZED STACKED BARS



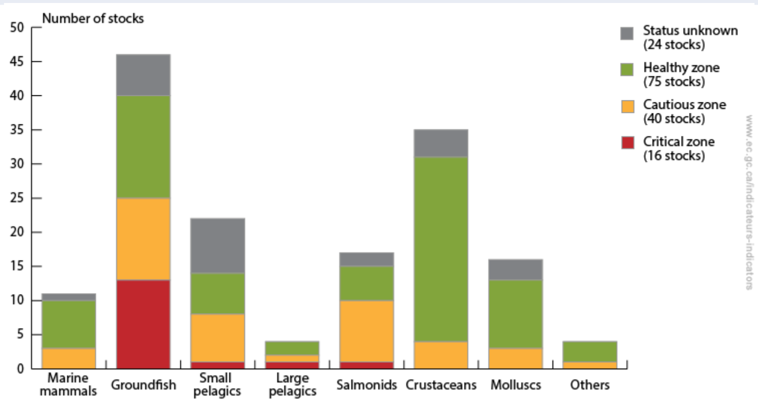
STACKED BARS



- ADVANTAGE
  - can compare totals and lowest level well
- DISADVANTAGE
  - upper levels of secondary key require comparison against non-aligned scale

STACKED BARS

DATA	MD table; one quantitative value attribute, two categorical key attributes
ENCODE	bar glyph: length-encoded subcomponents for each level of secondary key attribute separate bars by category of primary key
TASK	part-to-whole relationship, lookup values, find trends
SCALE	key attribute (main axis): dozens to hundreds of levels key attribute (stacked glyph axis): several to one dozen



STREAMGRAPH

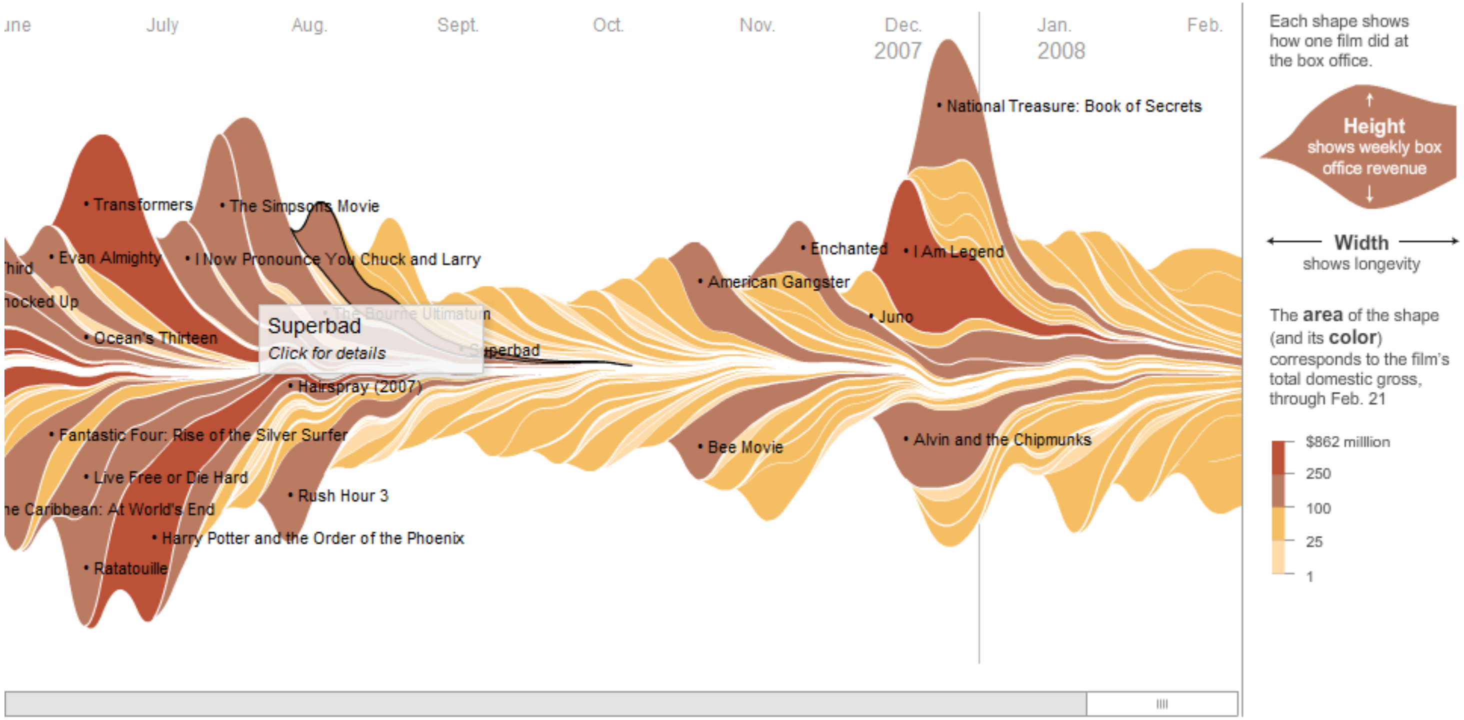
February 23, 2008

SIGN IN TO E-MAIL OR SAVE THIS | FEEDBACK

The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.

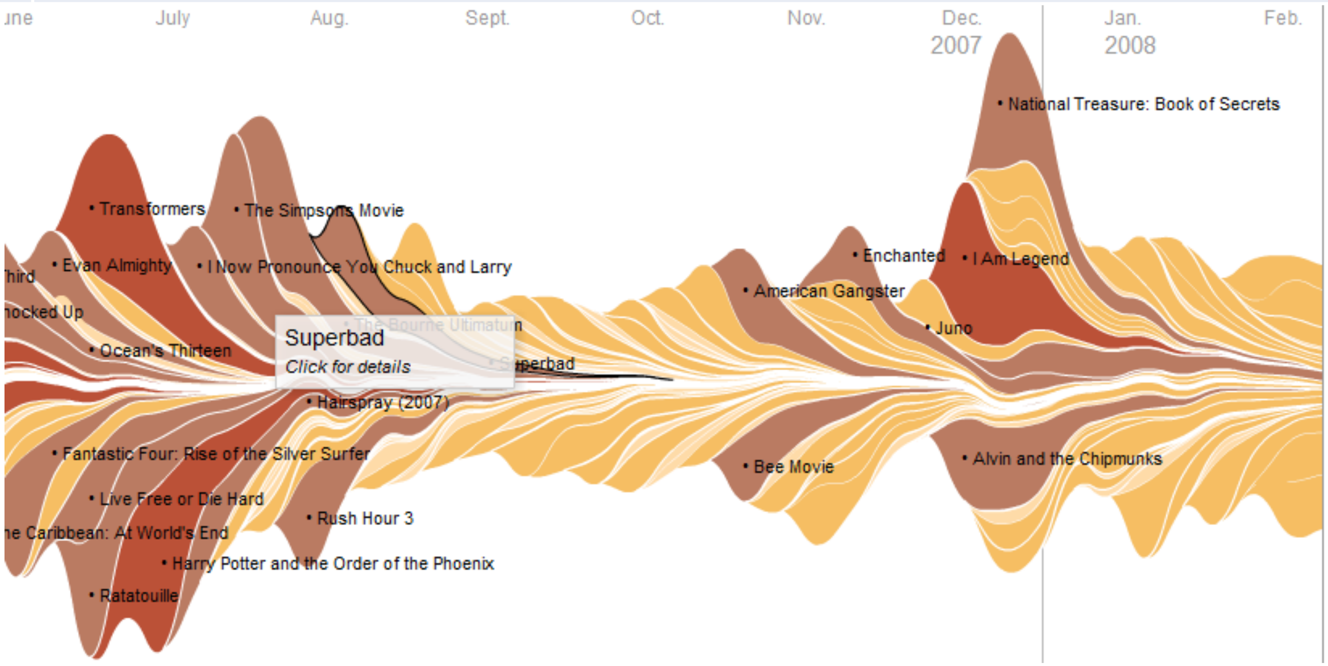
Find Movie Go



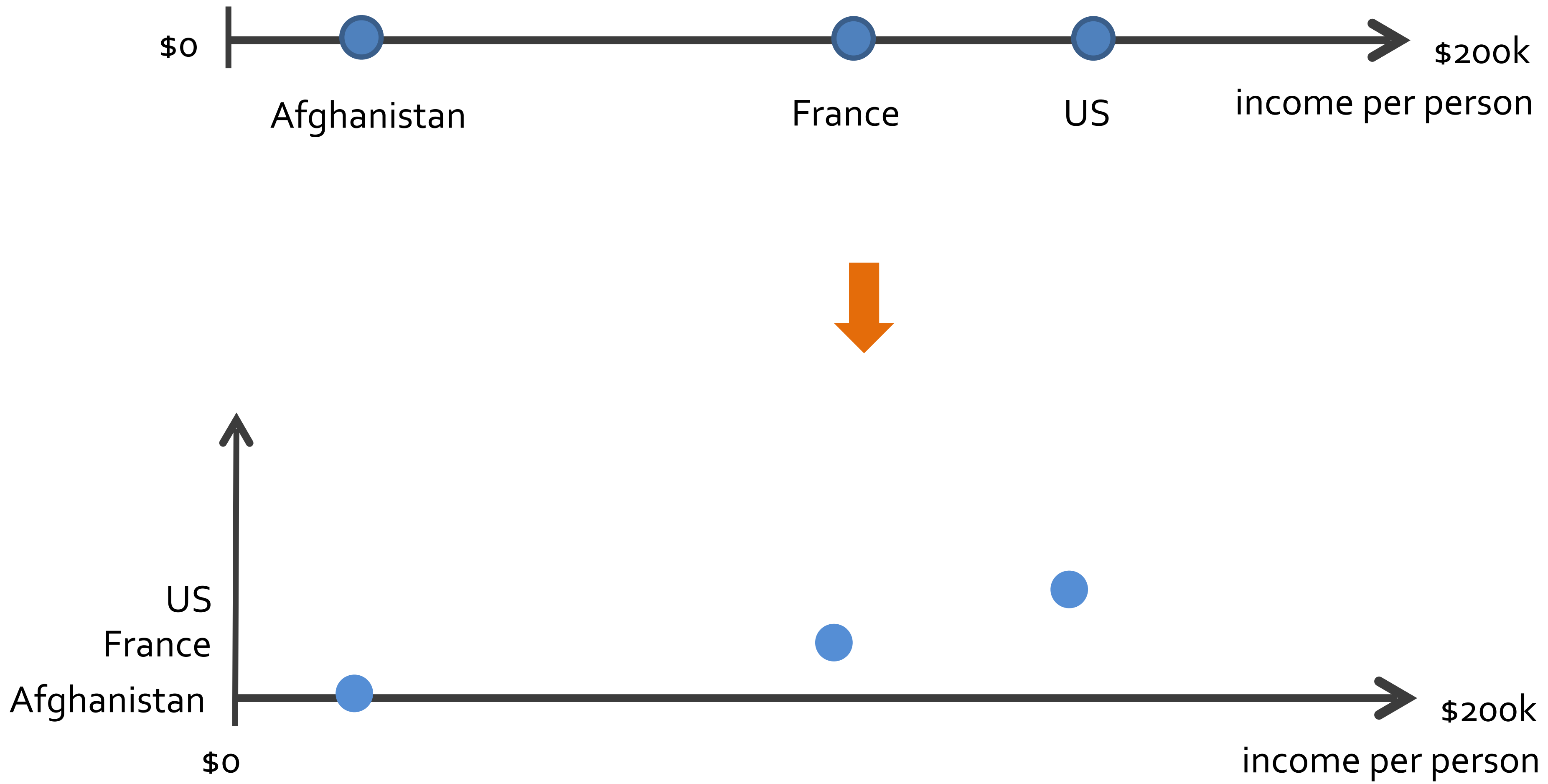
Sources: Baseline StudioSystems; Box Office Mojo

Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox

DATA	MD table; one quantitative value attribute (e.g. counts), one ordered key attribute (e.g. time), one categorical key attribute (e.g. film)
DERIVE	depth order of layers is derived from a quantitative attribute
ENCODE	use derived geometry to show layers across time, layer height encodes count
SCALE	key attributes (time, main axis): hundreds of time points key attributes (short axis): dozens to hundreds



DOT CHART/PLOT

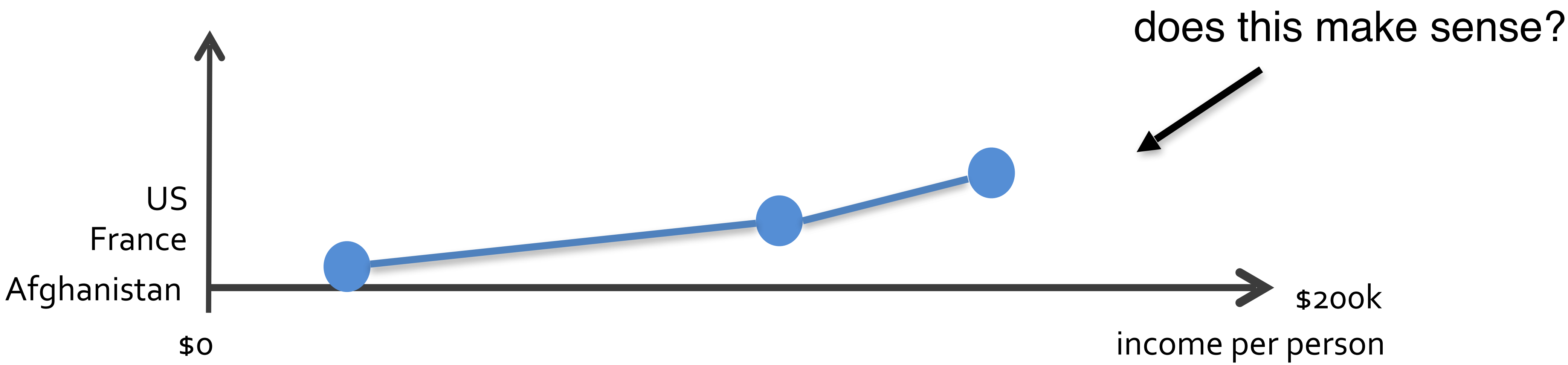


LINE CHART

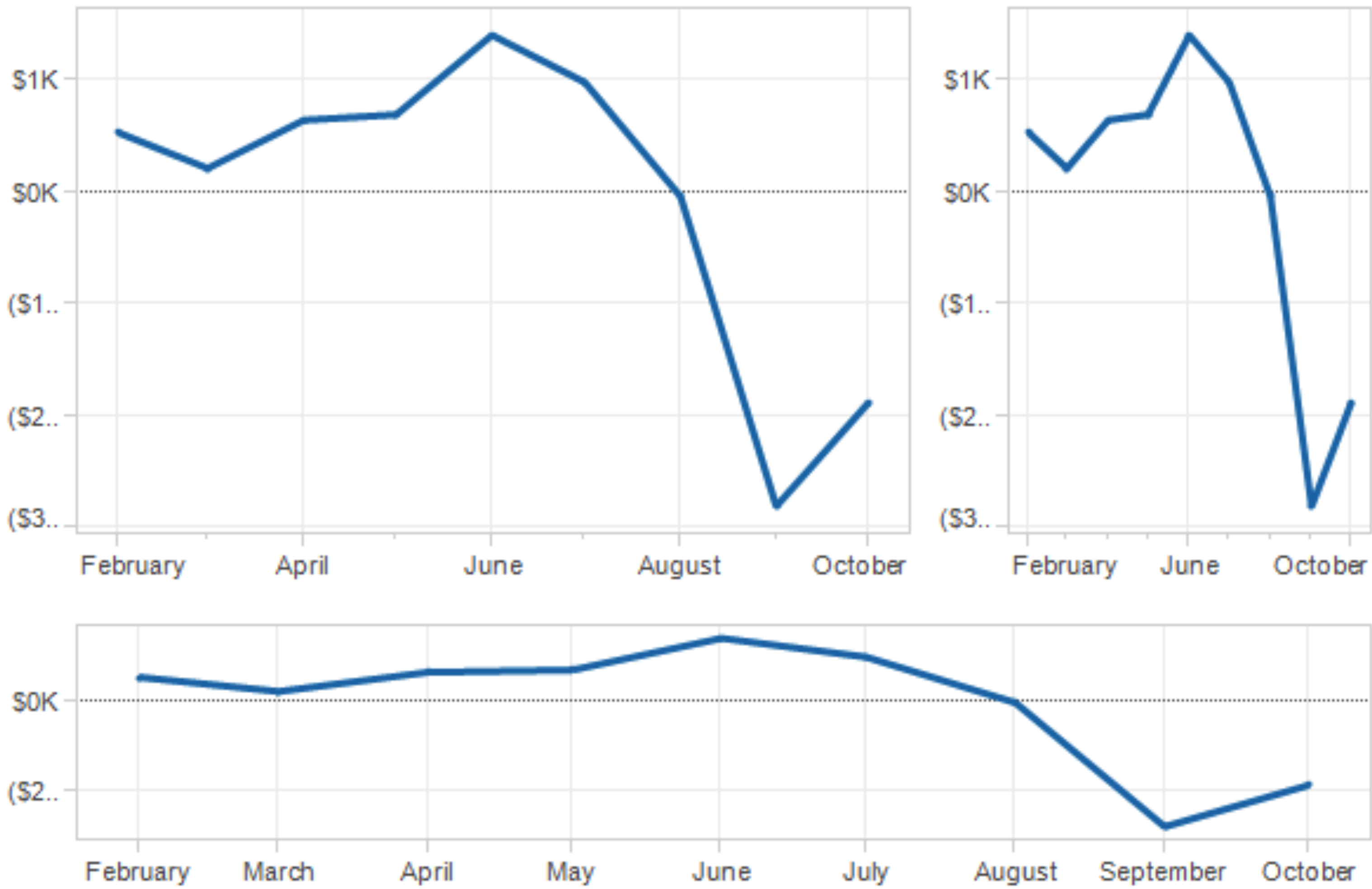
augment with line connection marks

emphasize the ordering and show trends

… should not be used with categorical keys!



ASPECT RATIO SELECTION

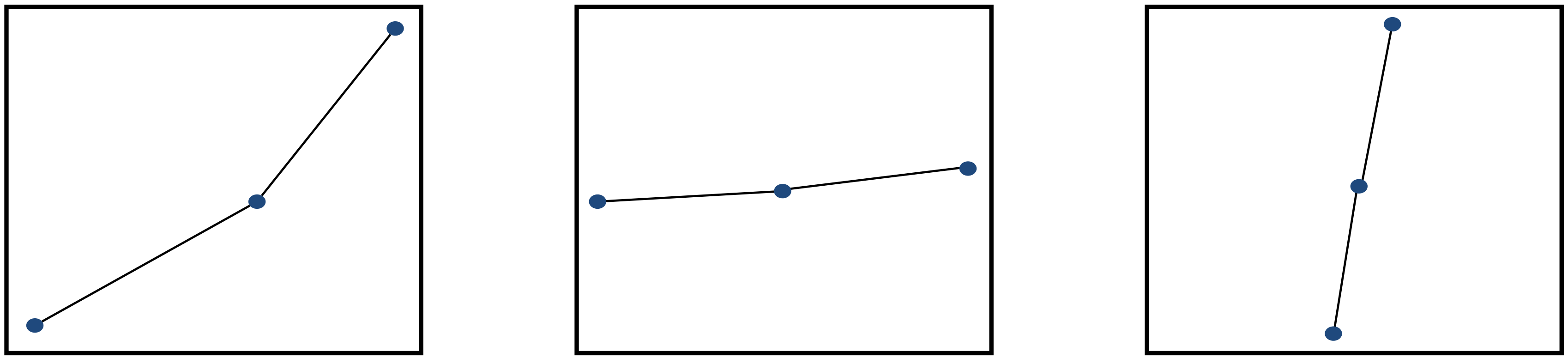


BANKING TO 45°

°°

[Cleveland]

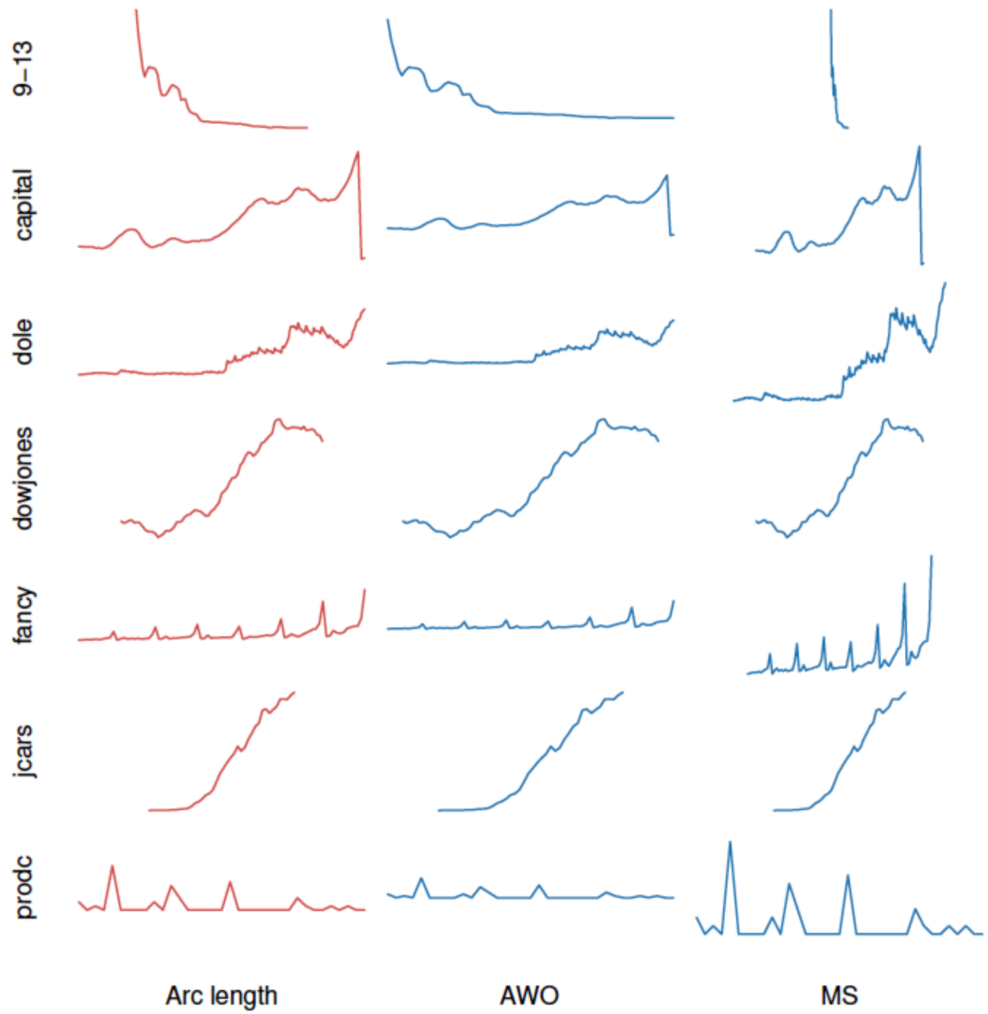
TO FACILITATE PERCEPTION OF TRENDS, MAXIMIZE THE DISCRIMINABILITY OF LINE SEGMENT ORIENTATIONS



TWO SEGMENTS ARE MAXIMALLY DISCRIMINABLE WHEN THEIR AVG ABSOLUTE ANGLE IS 45°

OPTIMIZE THE *ASPECT RATIO* TO BANK TO 45°

ALTERNATIVE METHODS



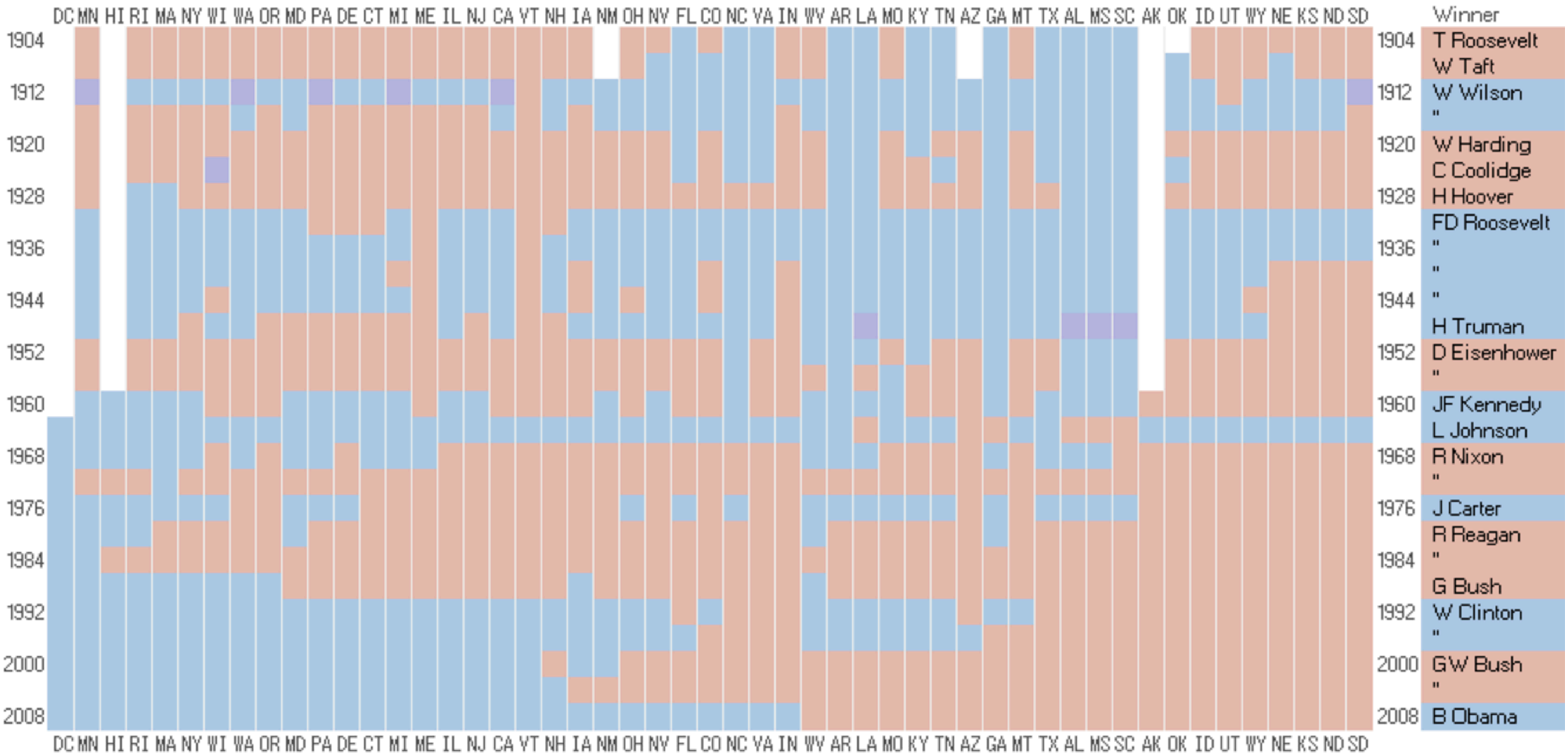
Practical advice:

CHOOSE AN ASPECT RATIO THAT EMPHASIZES THE IMPORTANT DETAILS FOR <u>YOUR</u> TASK

[TALBOT ET AL, 2011]

MATRIX ALIGNMENT

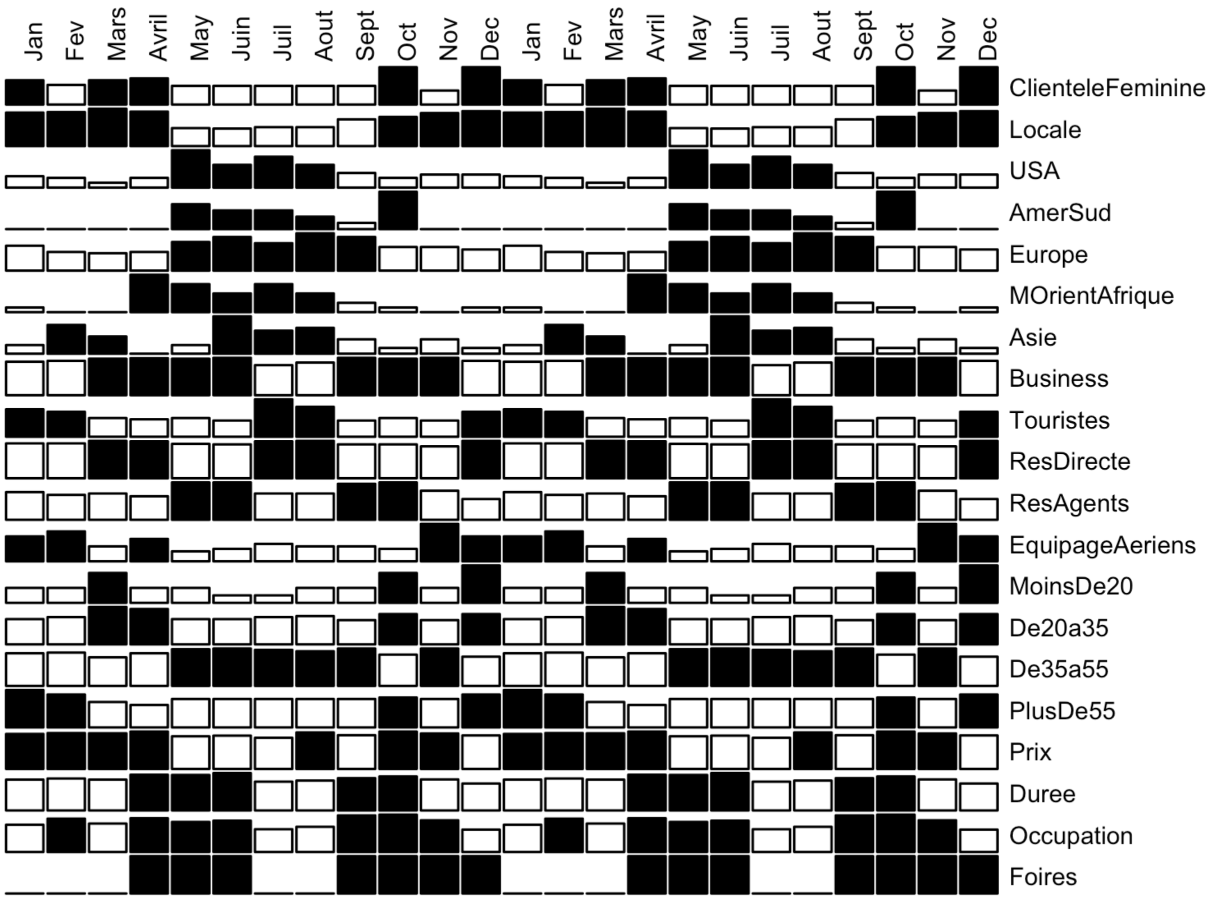
Two keys



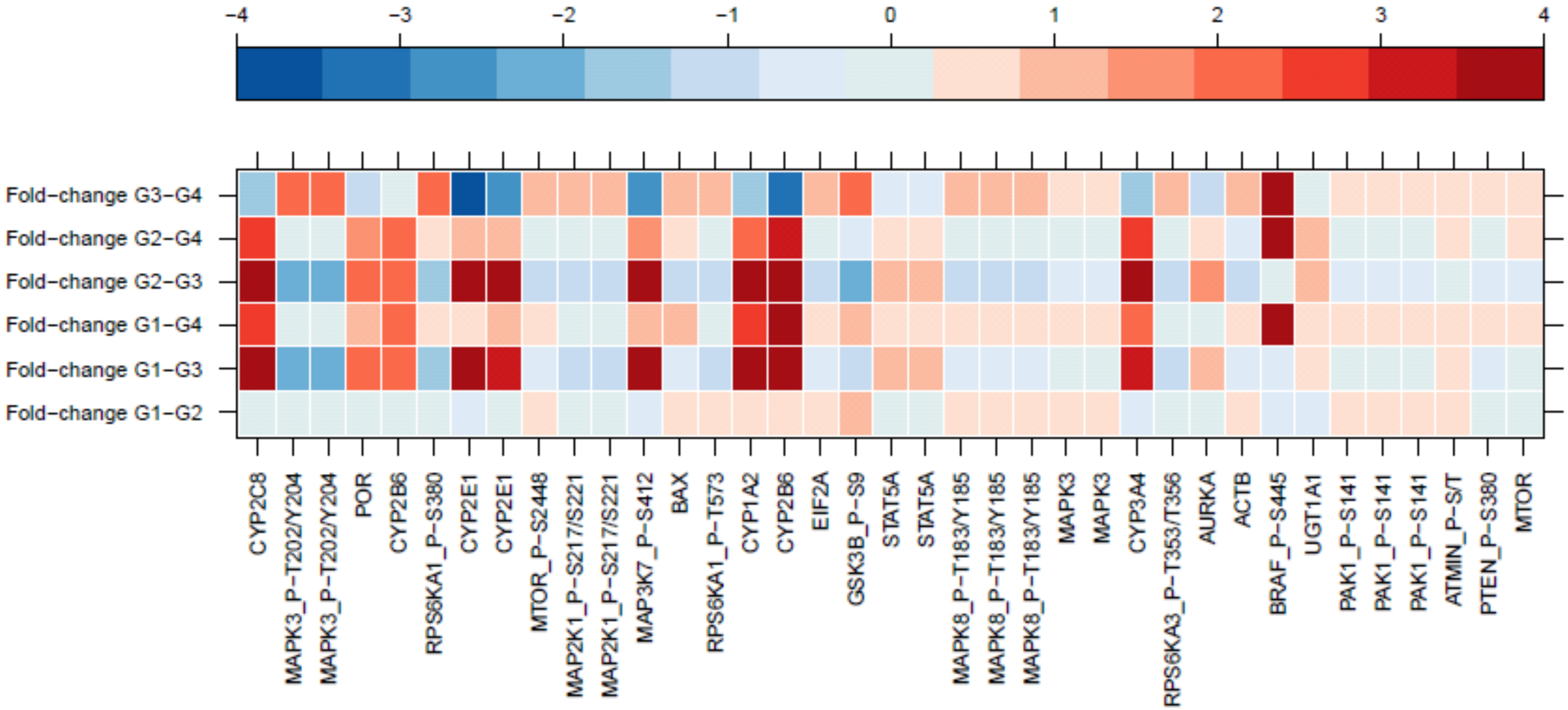
https://ldld.samizdat.cc/2016/tag/catalog/

HEATMAP

Hotel 2



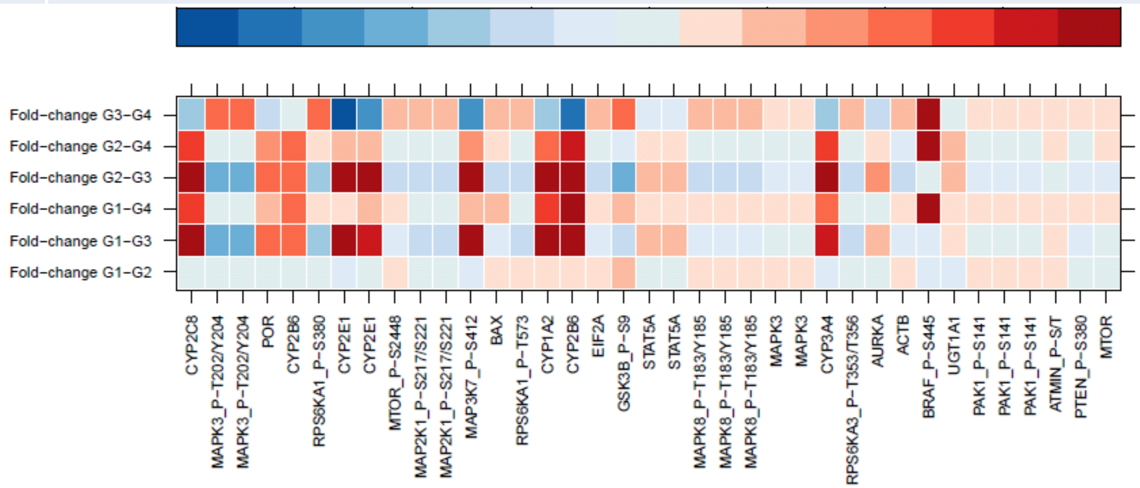
HEATMAP



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

HEATMAP

DATA	Table; two categorical key attributes, one quantitative value attribute
ENCODE	2D matrix alignment of area marks, e.g. with diverging color map
TASK	find clusters, outliers; summarize
SCALE	items: ~1 million (on 1000x1000px), categorical attribute levels: hundreds, quantitative attribute levels: 3-11

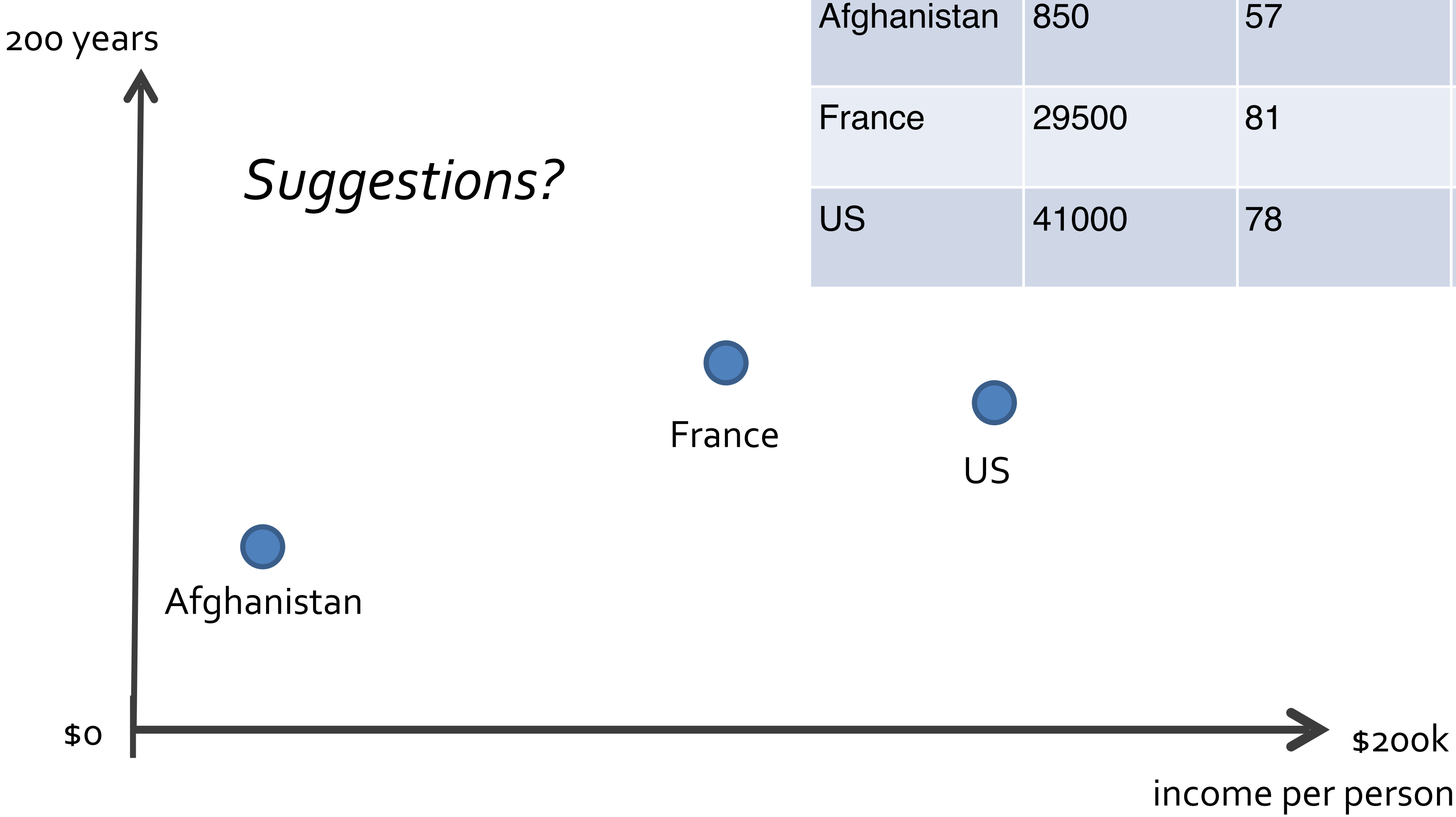


BACK TO OUR ORIGINAL EXAMPLE

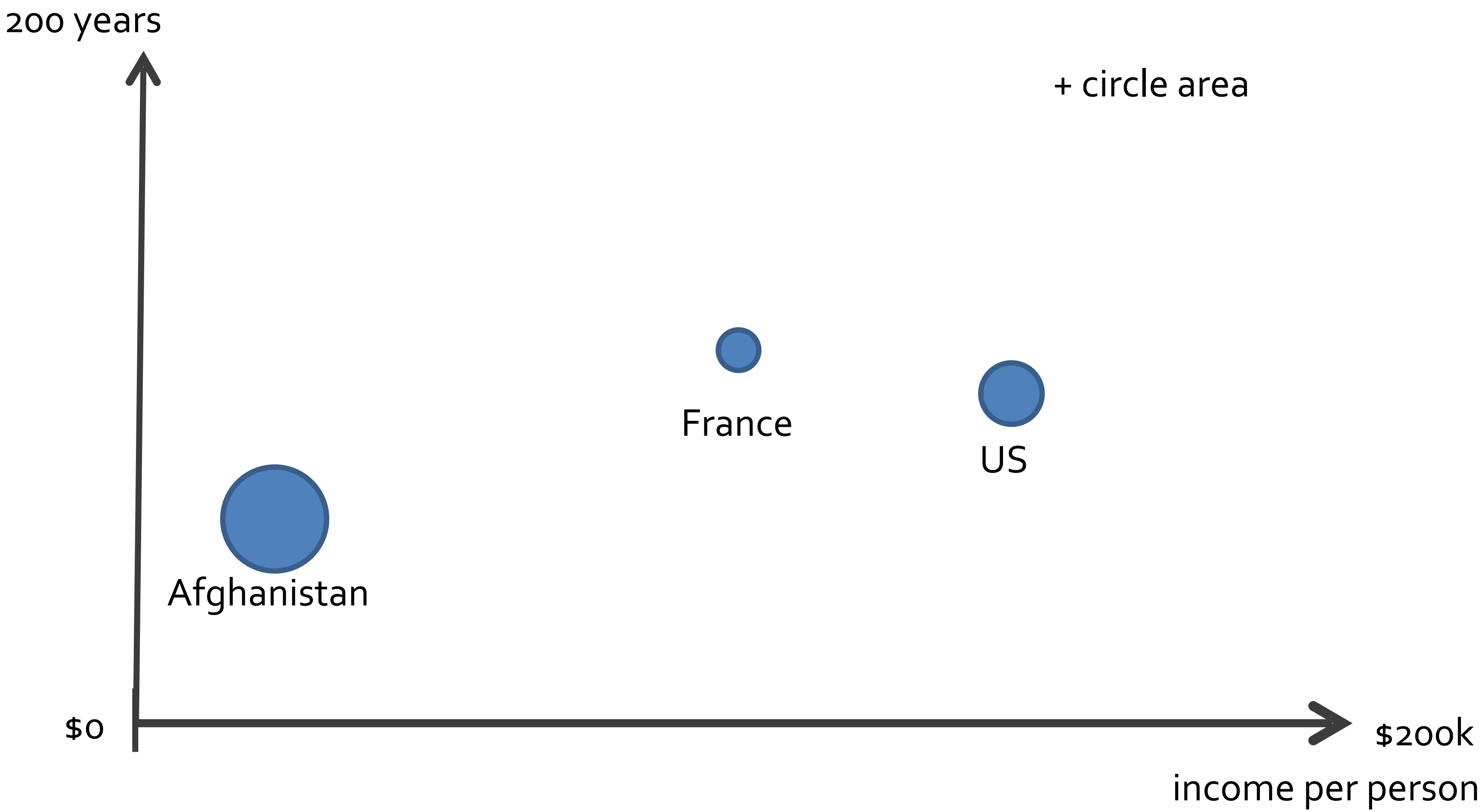
Country	Income per person	Life expectancy	Children per woman
Afghanistan	850	57	7.1
France	29500	81	1.9
US	41000	78	2.1

now with 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

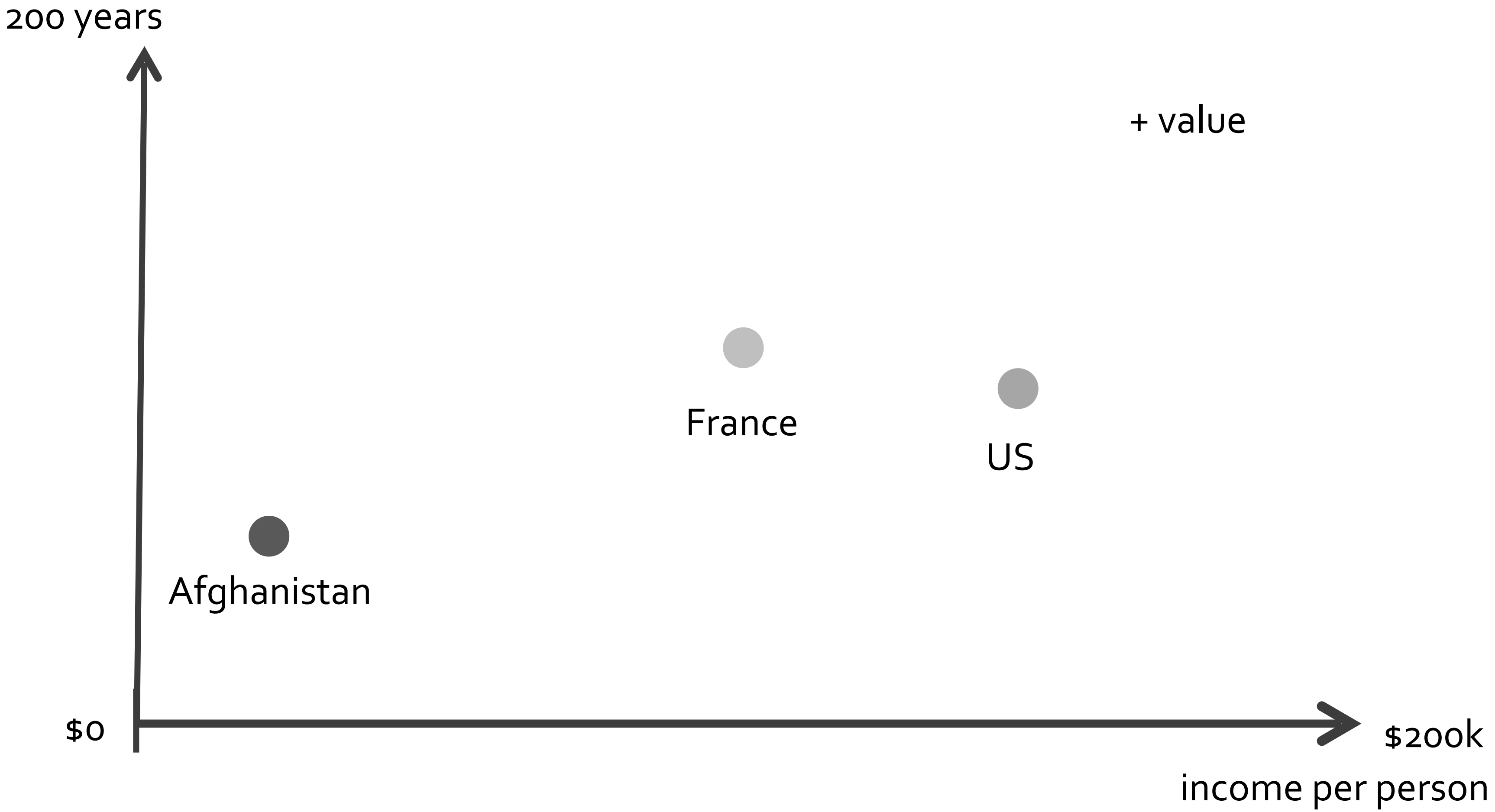


ADD ANOTHER VISUAL ENCODING



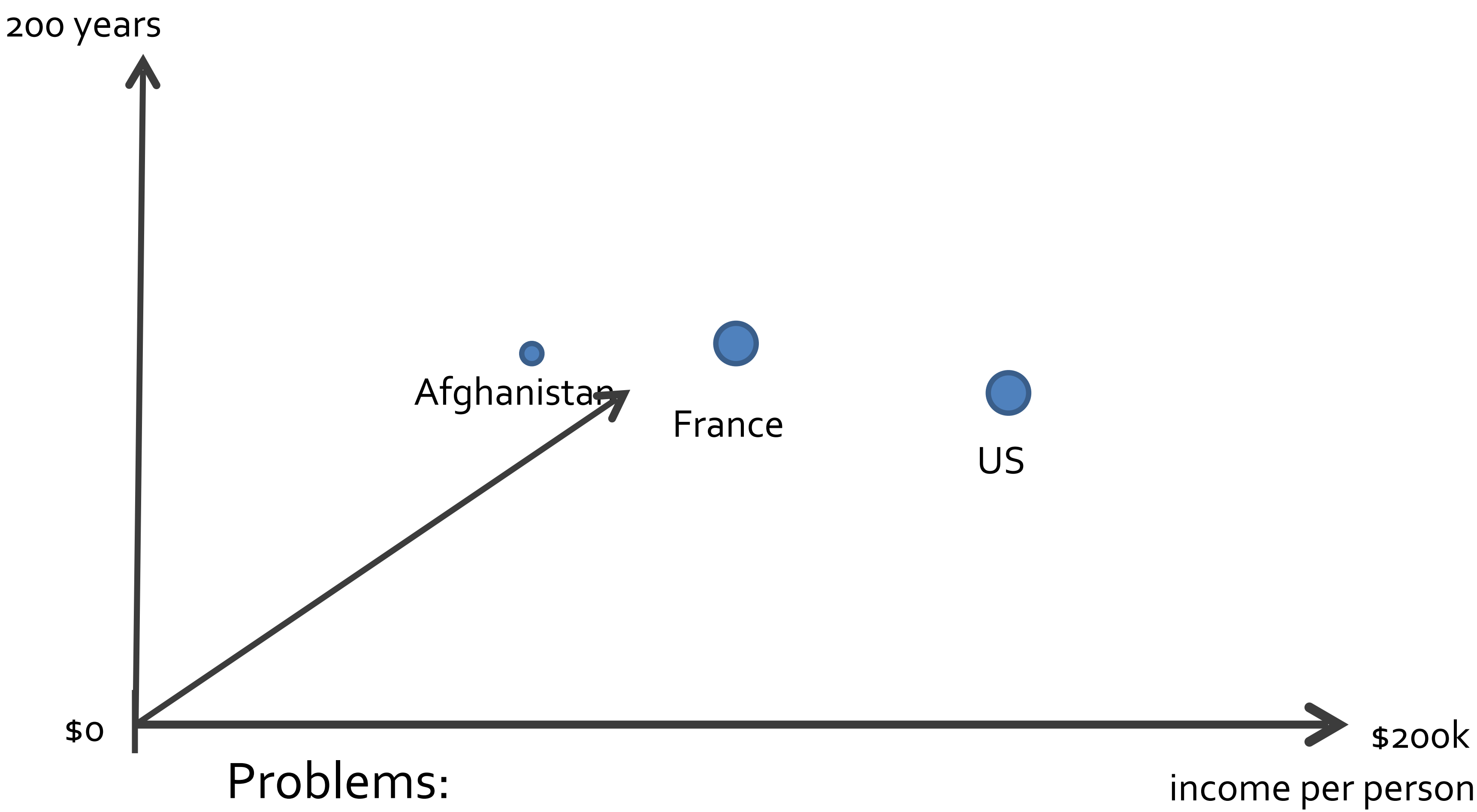
Problem:
Does not scale well to more attributes

ADD ANOTHER VISUAL ENCODING



Problem:
Does not scale well to more attributes

ADD AN AXIS

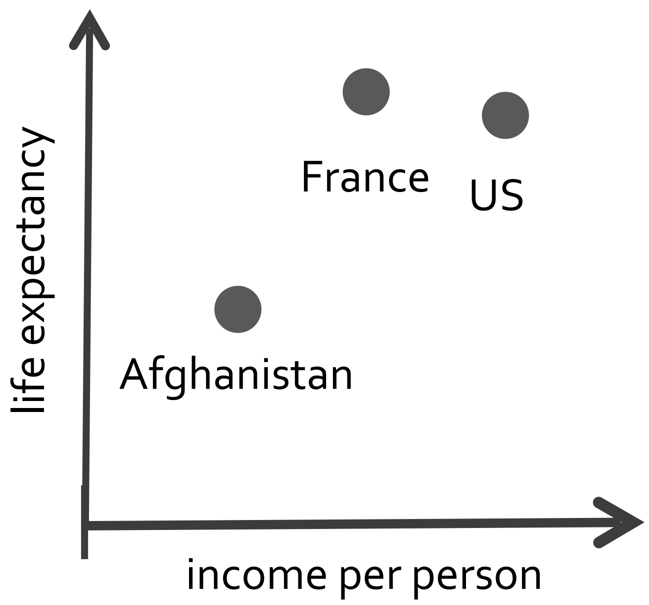


Problems:

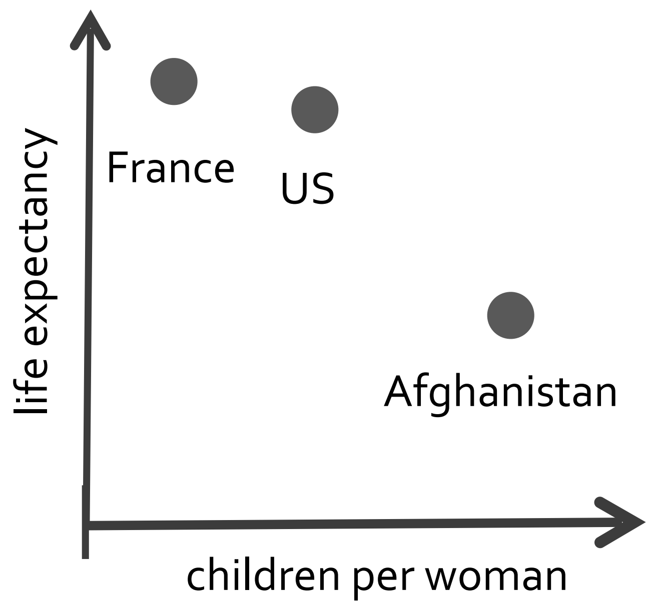
Occlusion, perspective distortion, does not scale

→ Not usually recommended

ADD AN AXIS



life expectancy
France
US
Afghanistan
income per person



life expectancy
France
US
Afghanistan
children per woman

SCATTERPLOT MATRIX

This idea scales relatively well

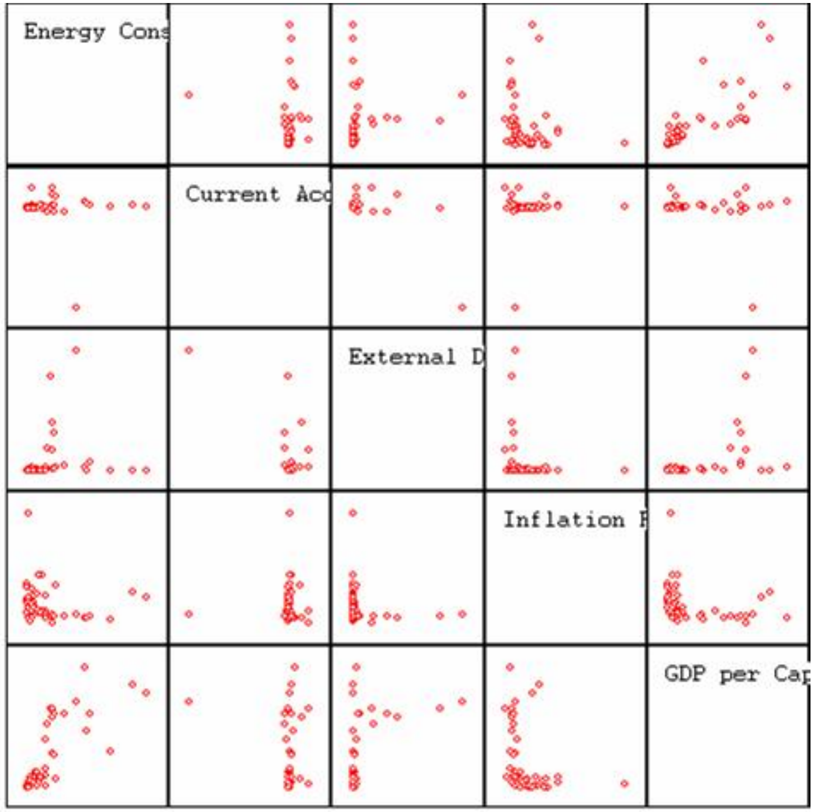


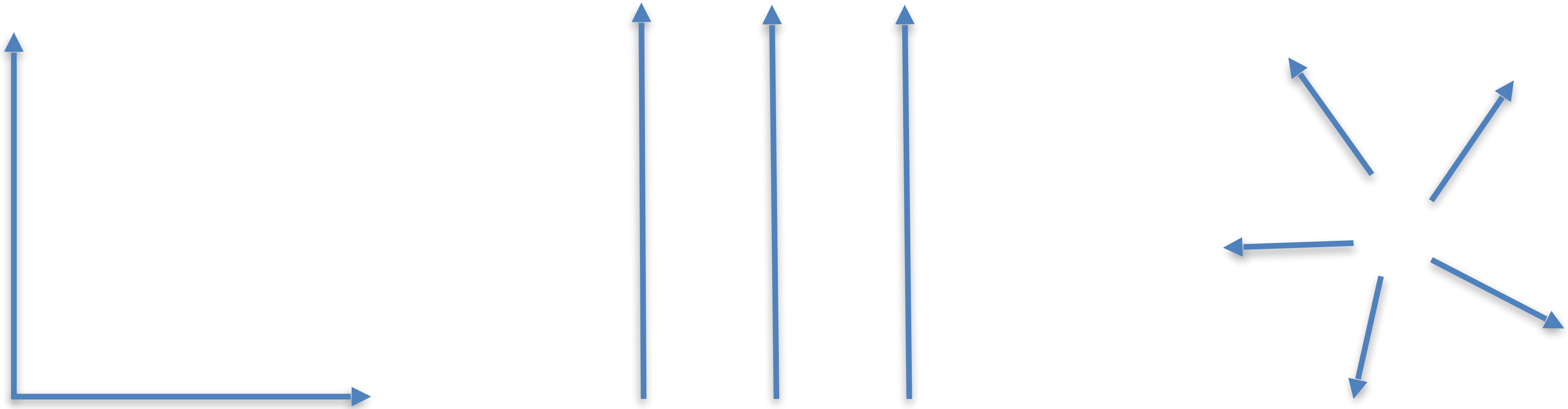
Image Source: Wikipedia

SCATTERPLOT MATRIX

DATA	Table
ENCODE	2D matrix alignment of area marks, e.g. with diverging color map
TASK	find clusters, outliers; summarize
SCALE	items: ~1 million (on 1000x1000px), categorical attribute levels: hundreds, quantitative attribute levels: 3-11

SPATIAL AXIS ORIENTATION

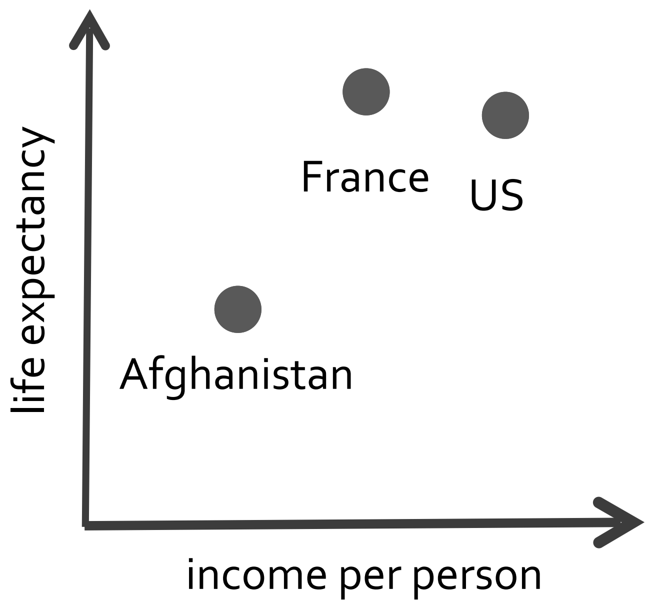
An additional design choice



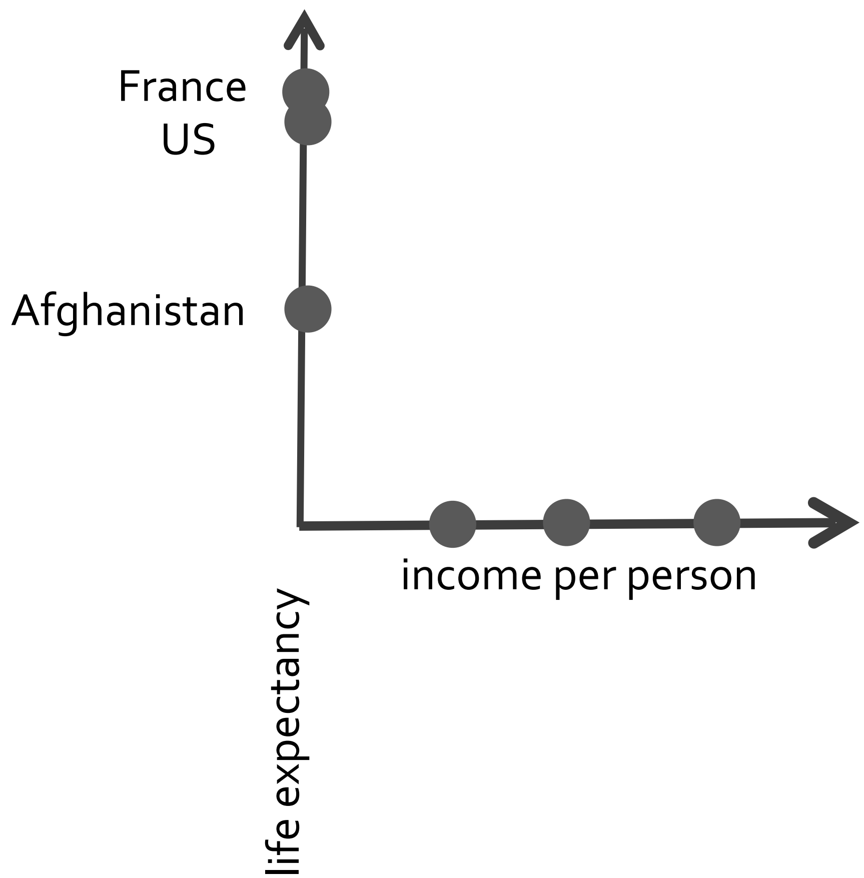
rectilinear | parallel | radial

parallel coordinates

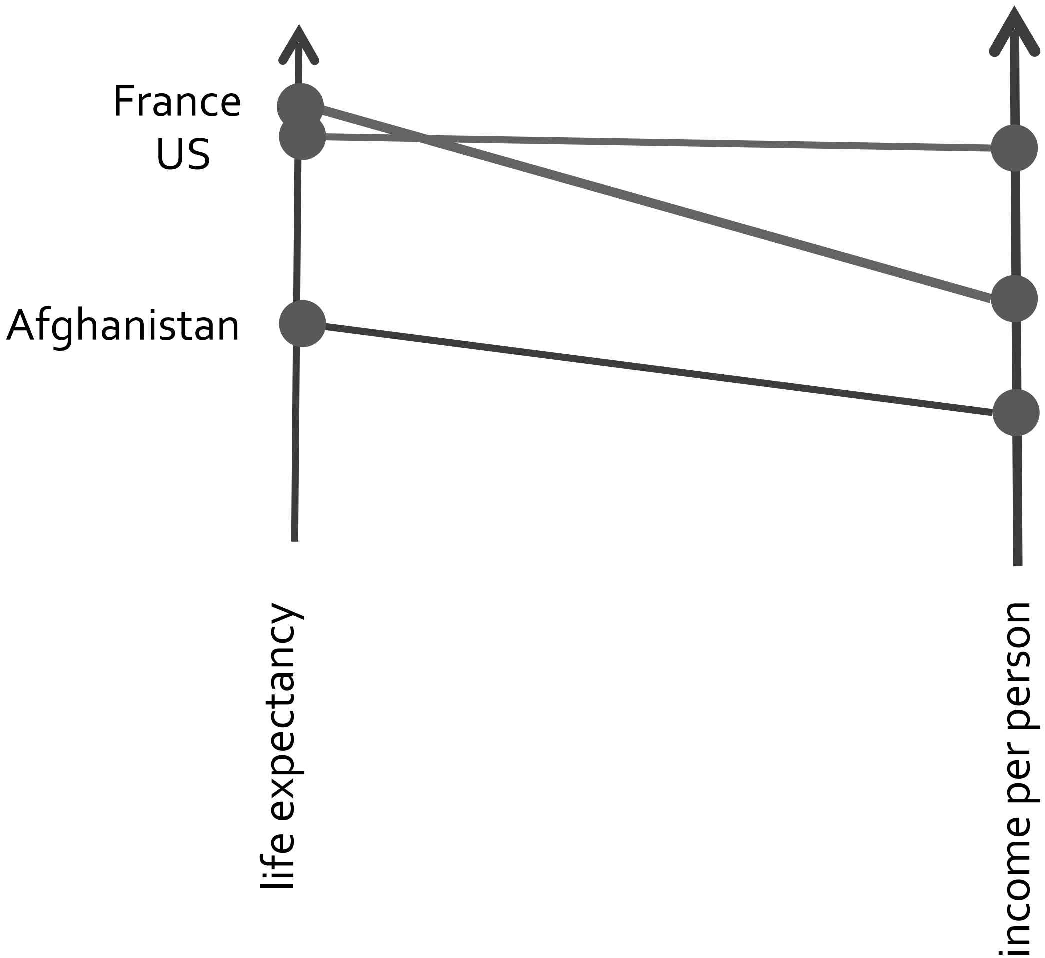
Back to our original example



Parallel Coordinates



parallel coordinates



- show correlations between neighboring axes

MULTIDIMENSIONAL DETECTIVE

Alfred Inselberg[*] Multidimensional Graphs Ltd[†]
&
Computer Science Department
Tel Aviv University, Israel
aiisreal@math.tau.ac.il

Abstract

The display of multivariate datasets in parallel coordinates, transforms the search for *relations* among the variables into a 2-D pattern recognition problem. This is the basis for the application to *Visual Data Mining*. The Knowledge Discovery process together with some general guidelines are illustrated on a dataset from the production of a VLSI chip. The special strength of parallel coordinates is in modeling **relations**. As an example, a simplified Economic Model is constructed with data from various economic sectors of a real country. The visual model shows the interelationship and dependencies between the sectors, circumstances where there is competition for the same resource, and feasible economic policies. Interactively, the model can be used to do trade-off analyses, discover sensitivities, do approximate optimization, monitor (as in a Process) and Decision Support.

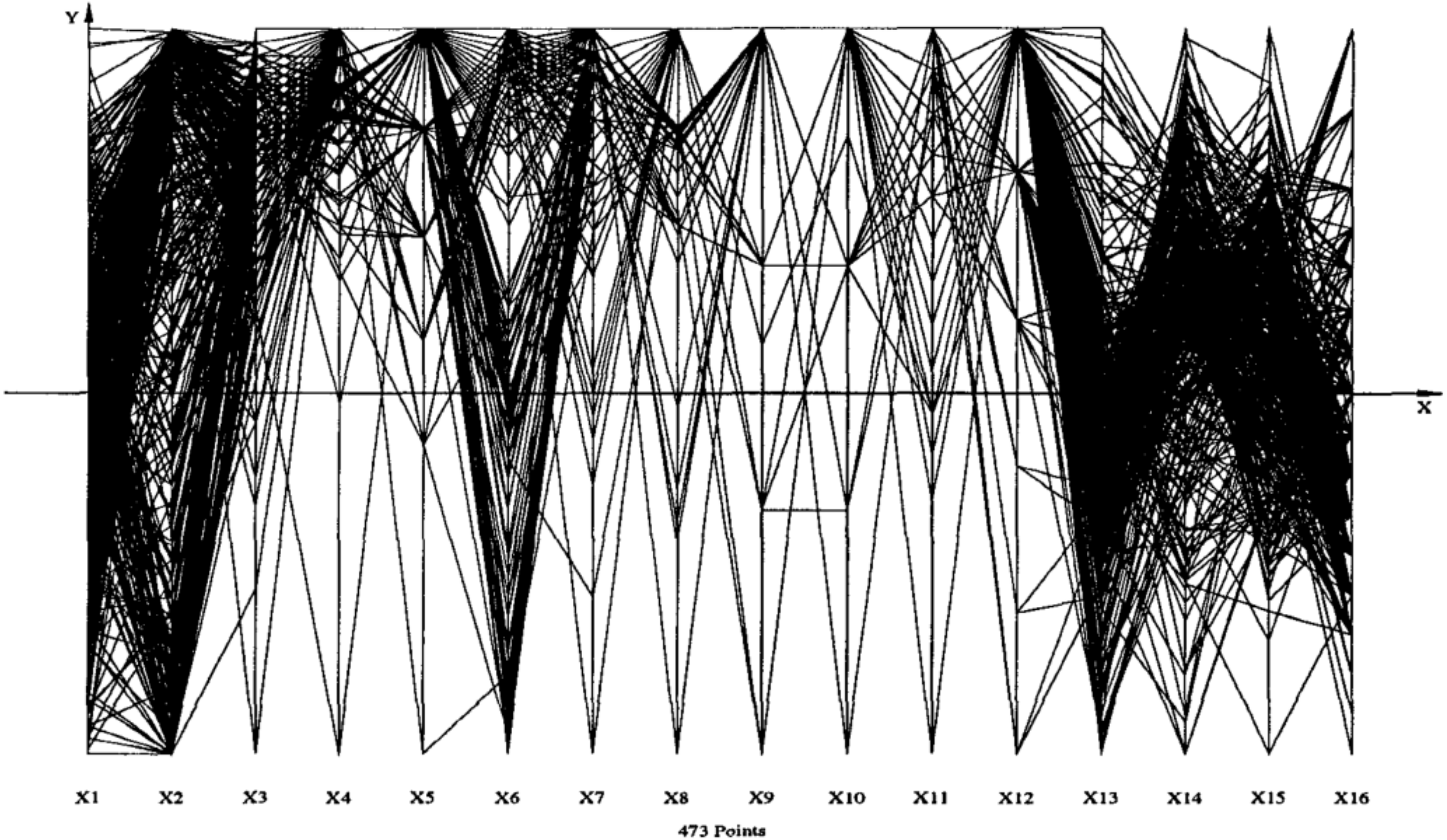
Introduction

In Geometry parallelism, which does not require a notion of angle, rather than orthogonality is the more fundamental concept. This, together with the fact that orthogonality "uses-up" the plane very fast, was the inspiration in 1959 for "Parallel" Coordinates. The systematic development began in 1977 [4]. The goals of the program were and still are (see [6] and [5] for short reviews) the visualization of multivariate/multidimensional problems without loss of information and having the properties:

1. Low representational complexity. Since the number of axes, N equals the number of dimensions (variables) the complexity is $O(N)$,
2. Works for any N,
3. Every variable is treated uniformly (unlike "Chernoff Faces" and various types of "glyphs"),
4. The displayed object can be recognized under projective transformations (i.e. rotation, translation, scaling, perspective),
5. The display easily/intuitively conveys information on the properties of the N-dimensional object it represents,
6. The methodology is based on rigorous mathematical and algorithmic results.

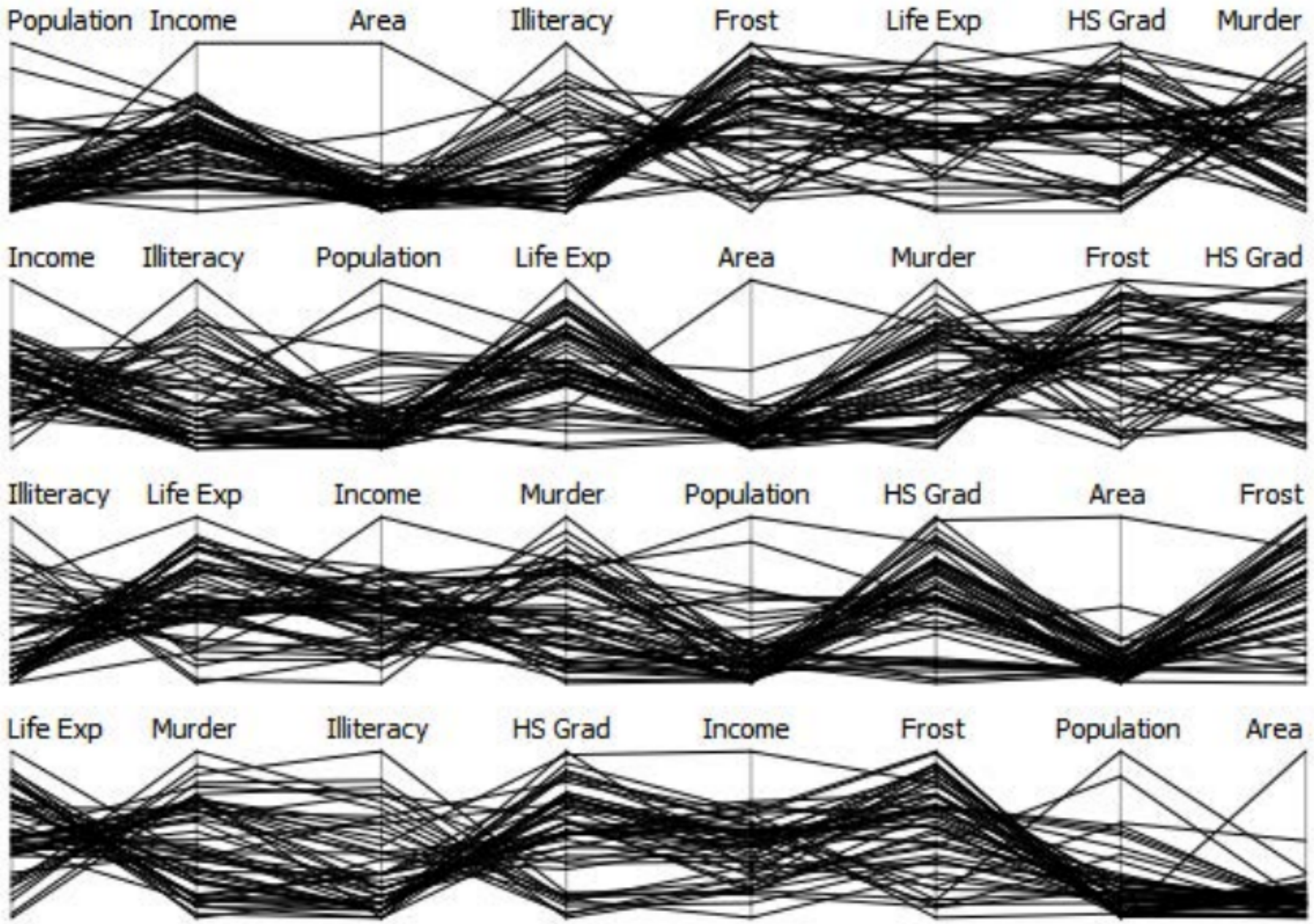
Parallel coordinates (abbr.||-coords) transform multivariate relations into 2-D patterns, a property that is well suited for Visual Data Mining.

* Senior Fellow San Diego SuperComputing Center
† 36A Yehuda Halevy Street, Raanana 43556, Israel



Original Example from Inselberg 1997

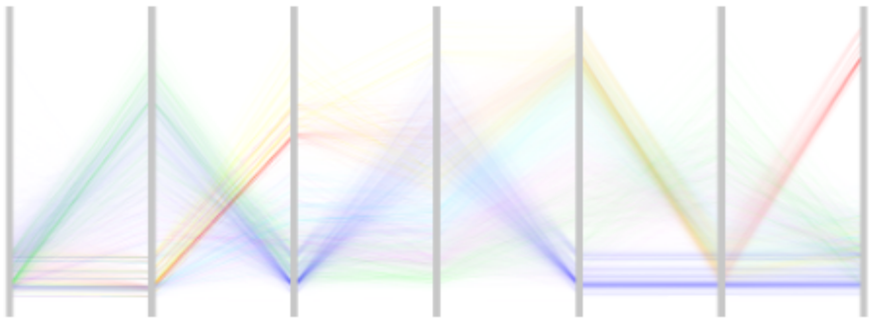
THE ORDER OF AXES MATTERS



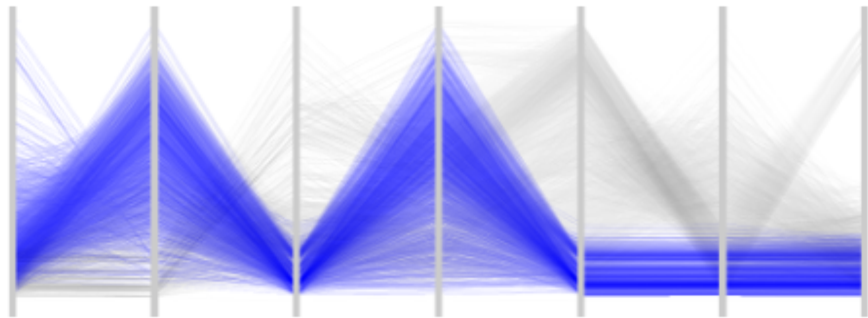
Eurographics 2013, STAR Report
J. Heinrich, D. Weiskopf

REDUCE CLUTTER - HIGHLIGHT CLUSTERS

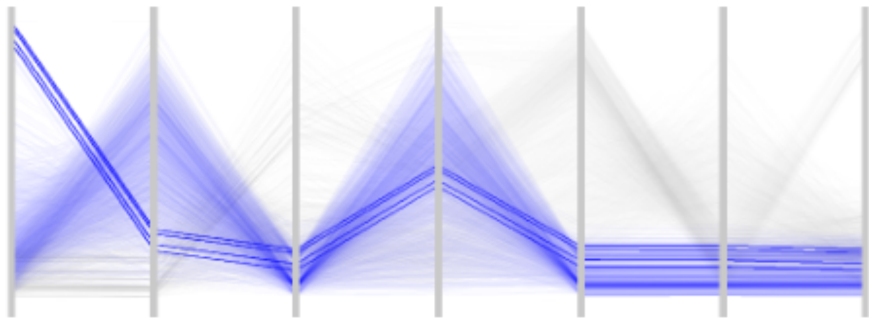
Lots of work on this. For example:



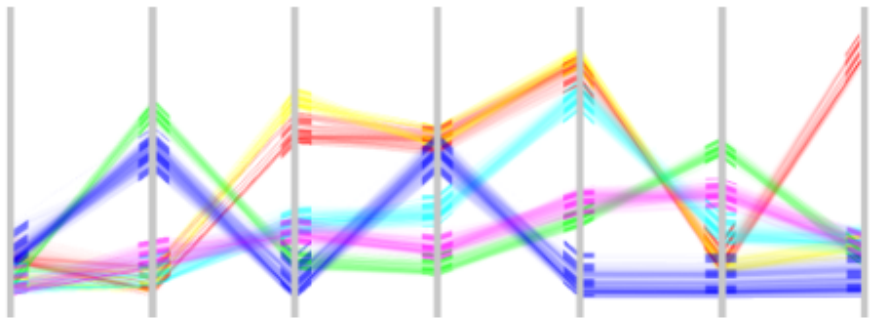
(a) A linear transfer function has been applied to the high-precision texture in order to prevent cluttering and to provide overview of the data.



(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.

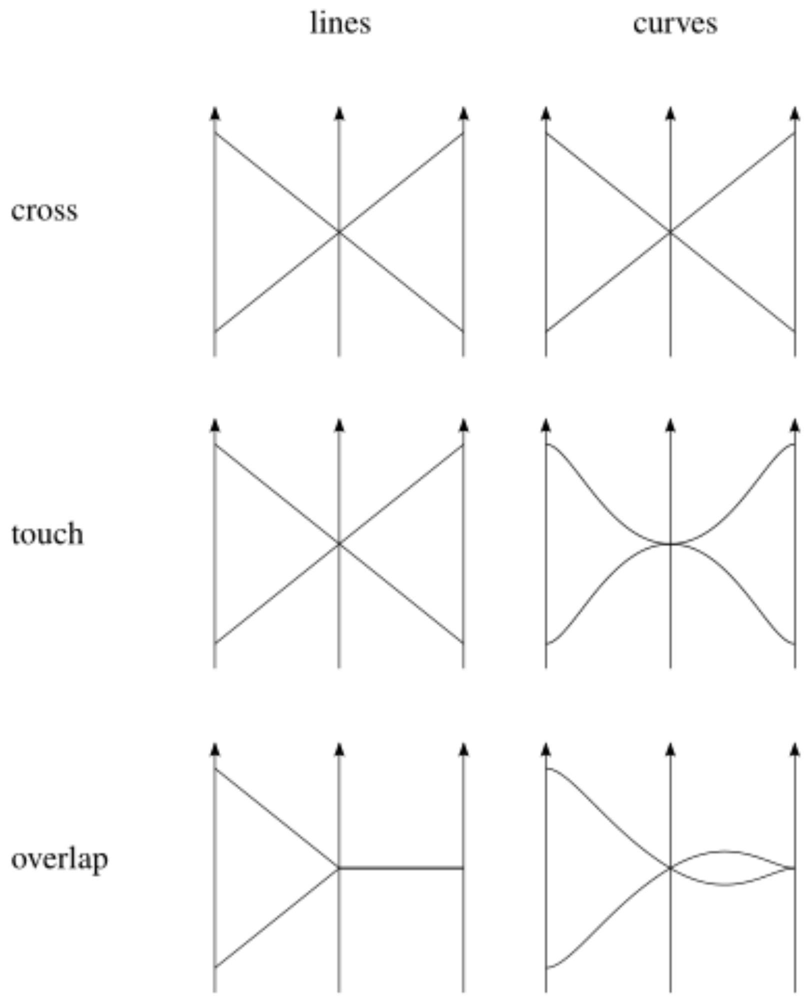


(d) A complementary view of the clusters with uniform bands. 'Feature animation' presents statistics about the clusters and acts as a guidance.

Revealing Structure within Clustered Parallel Coordinates Displays, InfoVis 2005

HOW TO DRAW THE LINES

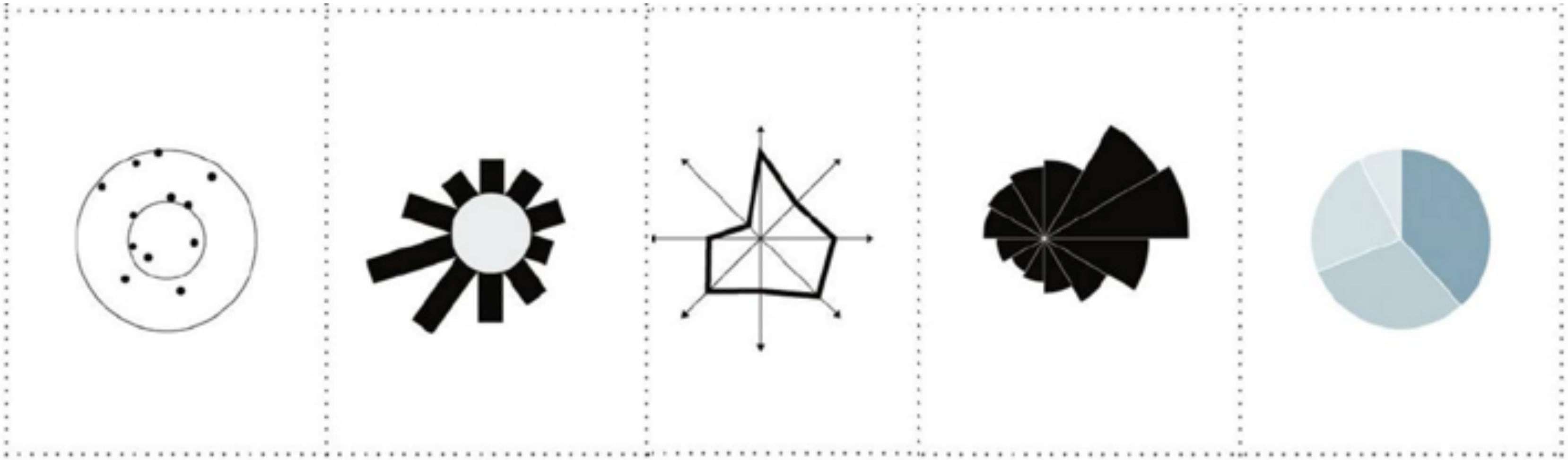
Goal: avoid ambiguity



Eurographics 2013, STAR Report
J. Heinrich, D. Weiskopf

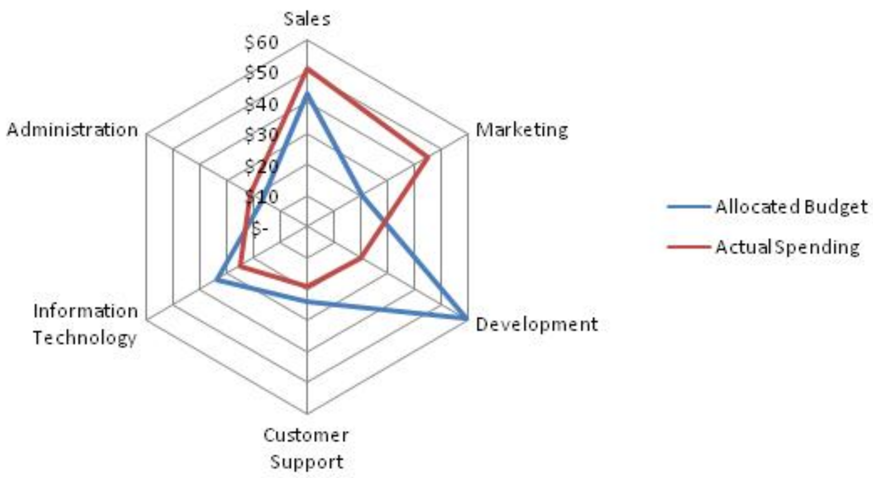
RADIAL AXES

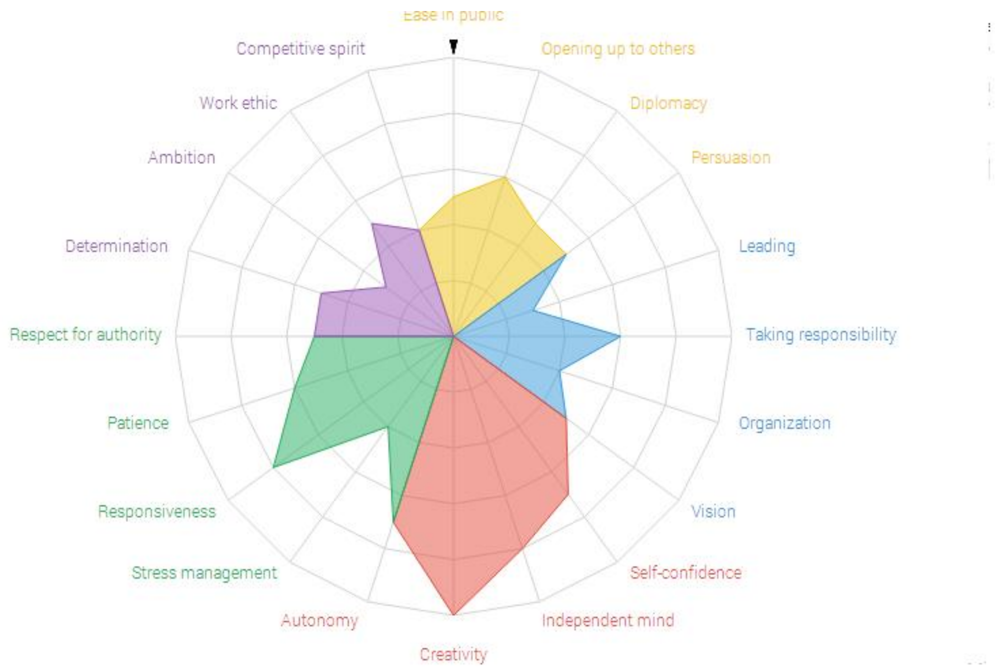
Polar



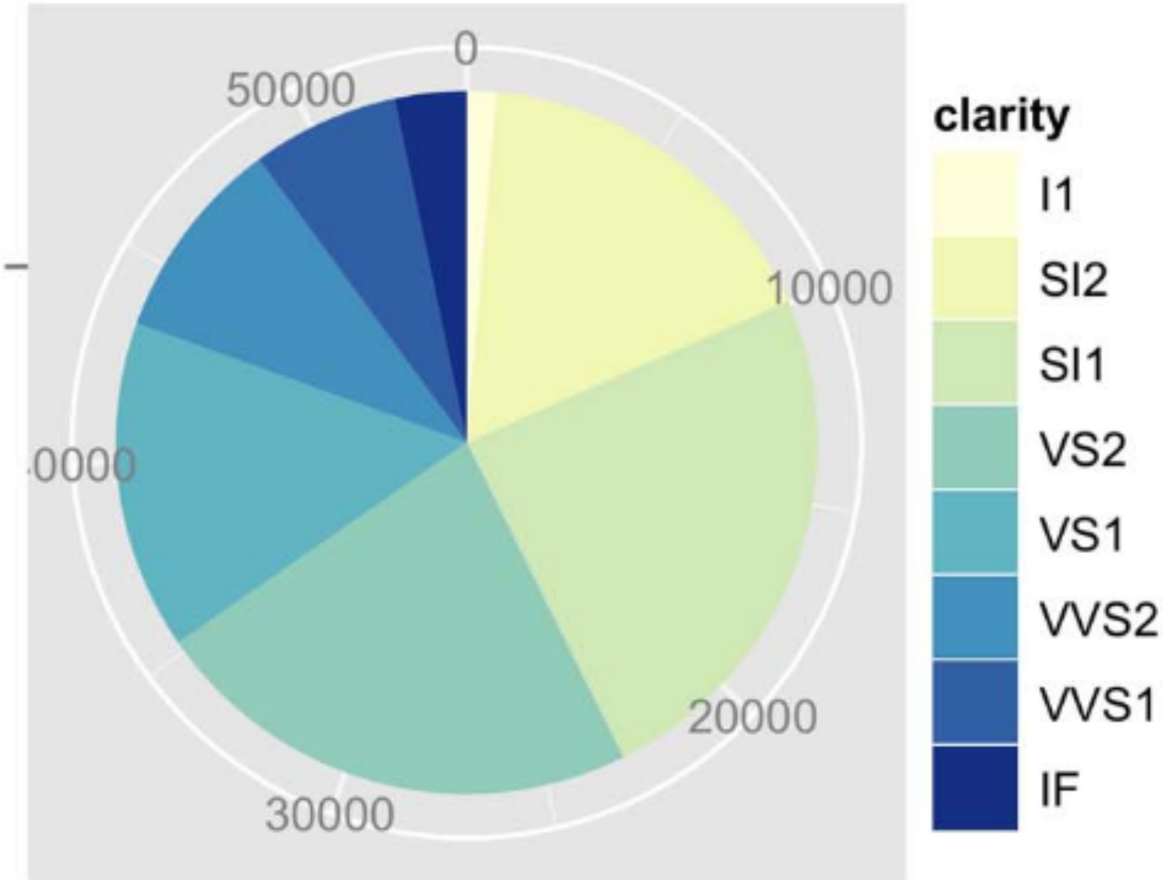
EXAMPLE: STAR PLOT

- = radial line chart

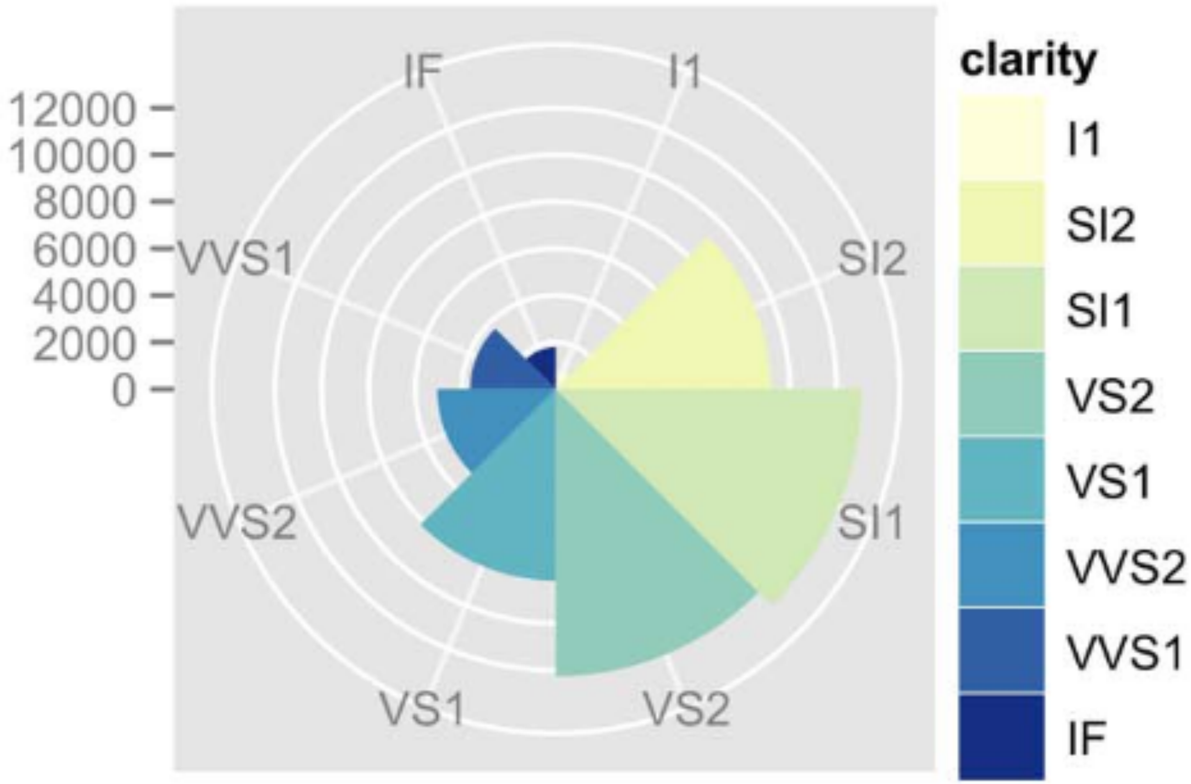




PIE CHARTS


0
50000
10000
0000
20000
30000
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF

POLAR AREA CHARTS



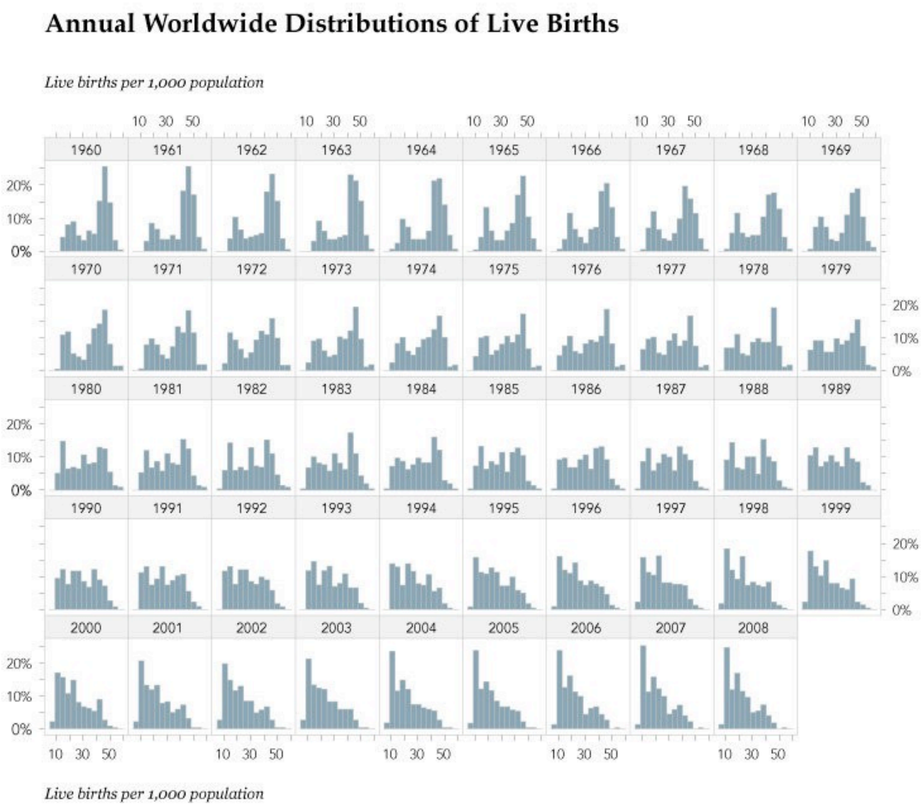
SPATIAL LAYOUT DENSITY

DATA DENSITY

MAXIMIZE THE RATIO OF:

$$\frac{\text{(}\textbf{NUMBER OF ENTRIES}\text{ IN DATA)}}{\text{(}\textbf{AREA}\text{ OF THE GRAPHIC)}}$$

DATA DENSITY – SHRINK THE GRAPHICS



“SMALL MULTIPLES”

DATA DENSITY – SHRINK THE GRAPHICS

Placed in the relevant context, a single number gains meaning. Thus the most recent measurement of glucose should be compared with earlier measurements for the patient. This data-line shows the path of the last 80 readings of glucose:

 glucose 6.6

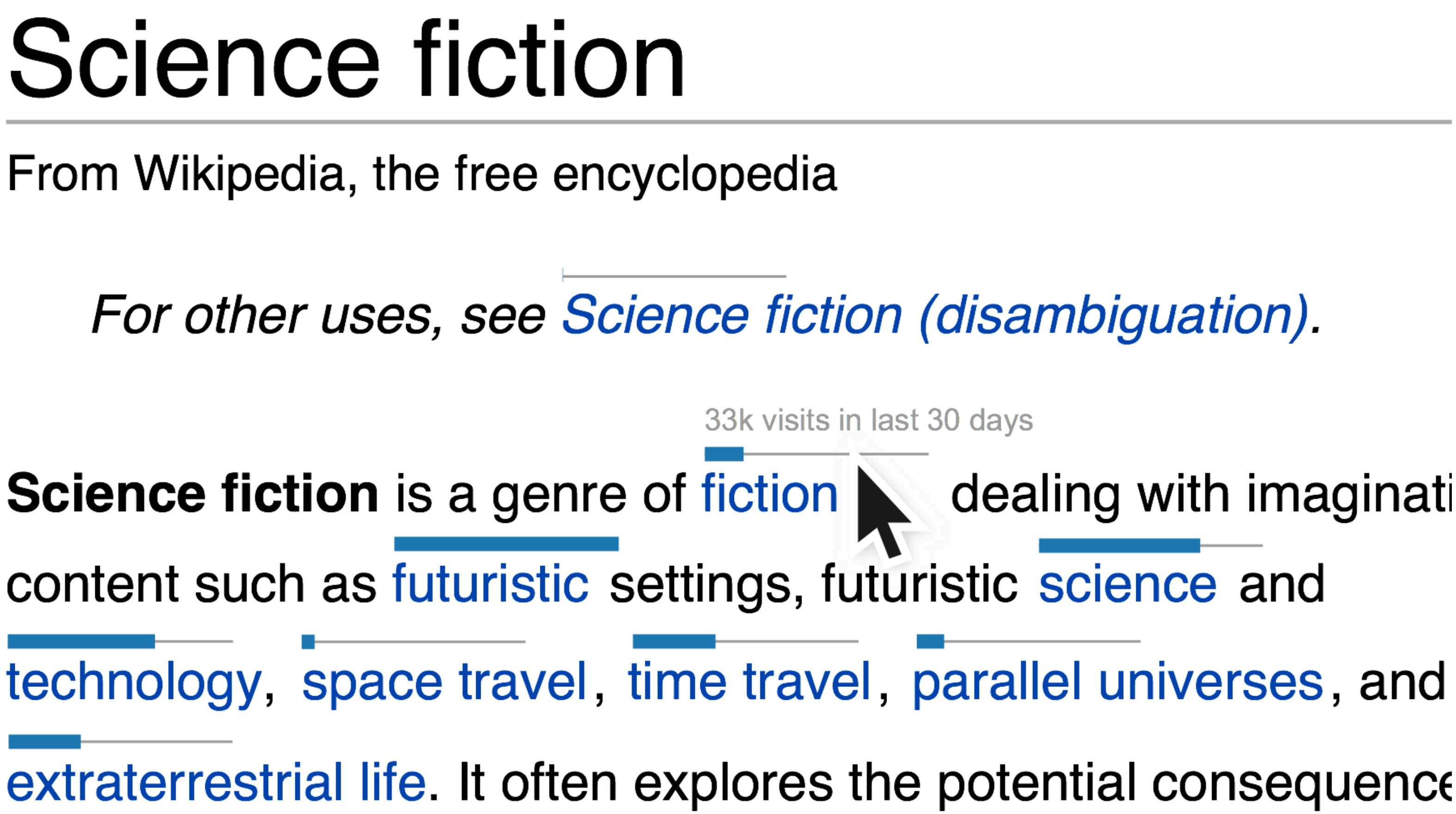
Lacking a scale of measurement, this free-floating line is dequantified. At least we do know the value of the line's right-most data point, which corresponds to the most recent value of glucose, the number recorded at far right. Both representations of the most recent reading are tied together with a color accent:

 glucose 6.6

Some useful context is provided by showing the *normal range* of glucose, here as a gray band. Compared to normal limits, readings above the band horizon are elevated, those below reduced:

 glucose 6.6 or glucose  6.6

SPARKLINES & WORD-SCALE VIS



SPARKLINES & WORD-SCALE VIS

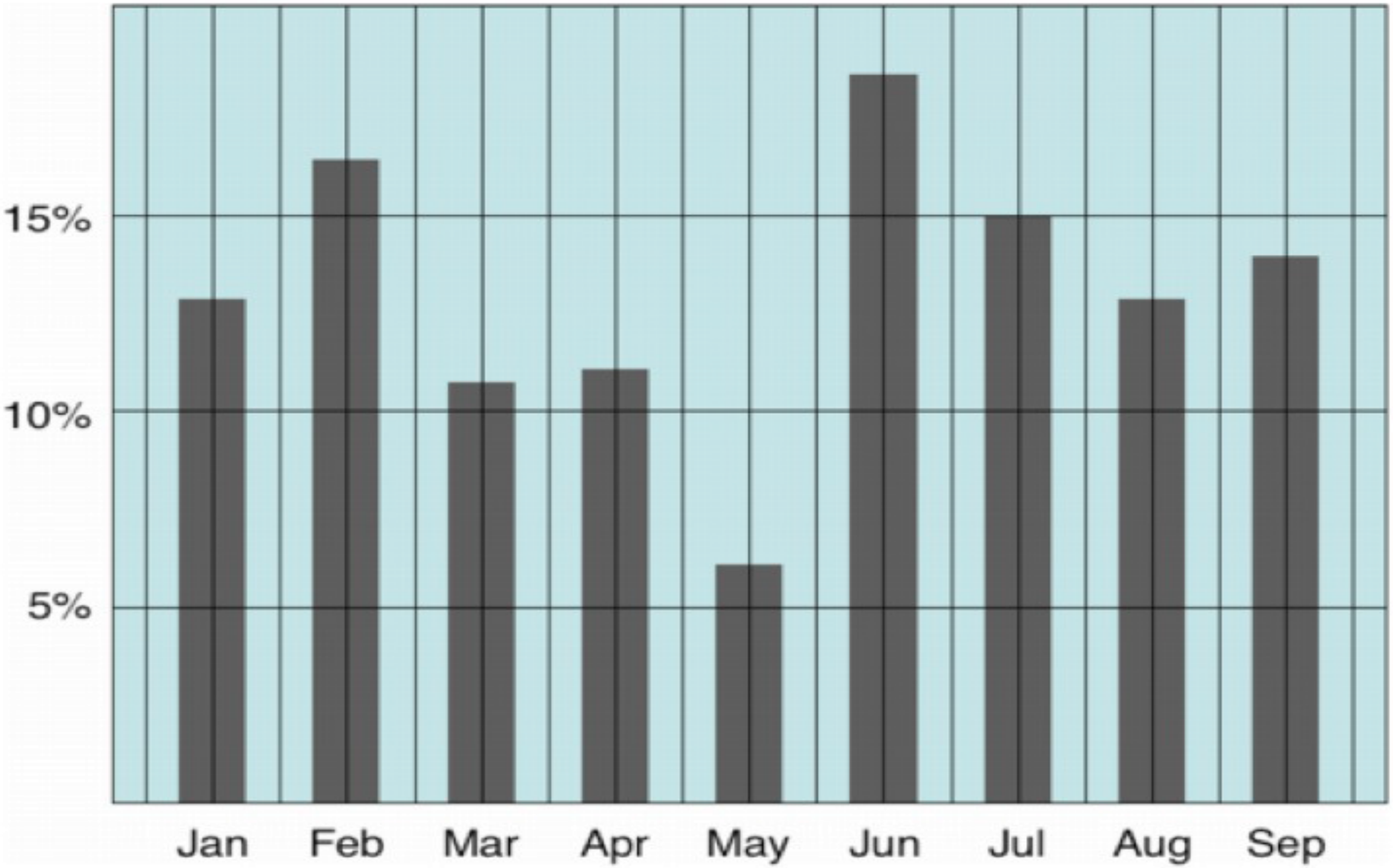
Gonzalo Higuaín slides a cross in from the right  and Ronaldo, at the front post, shoots off target.

DATA-INK RATIO

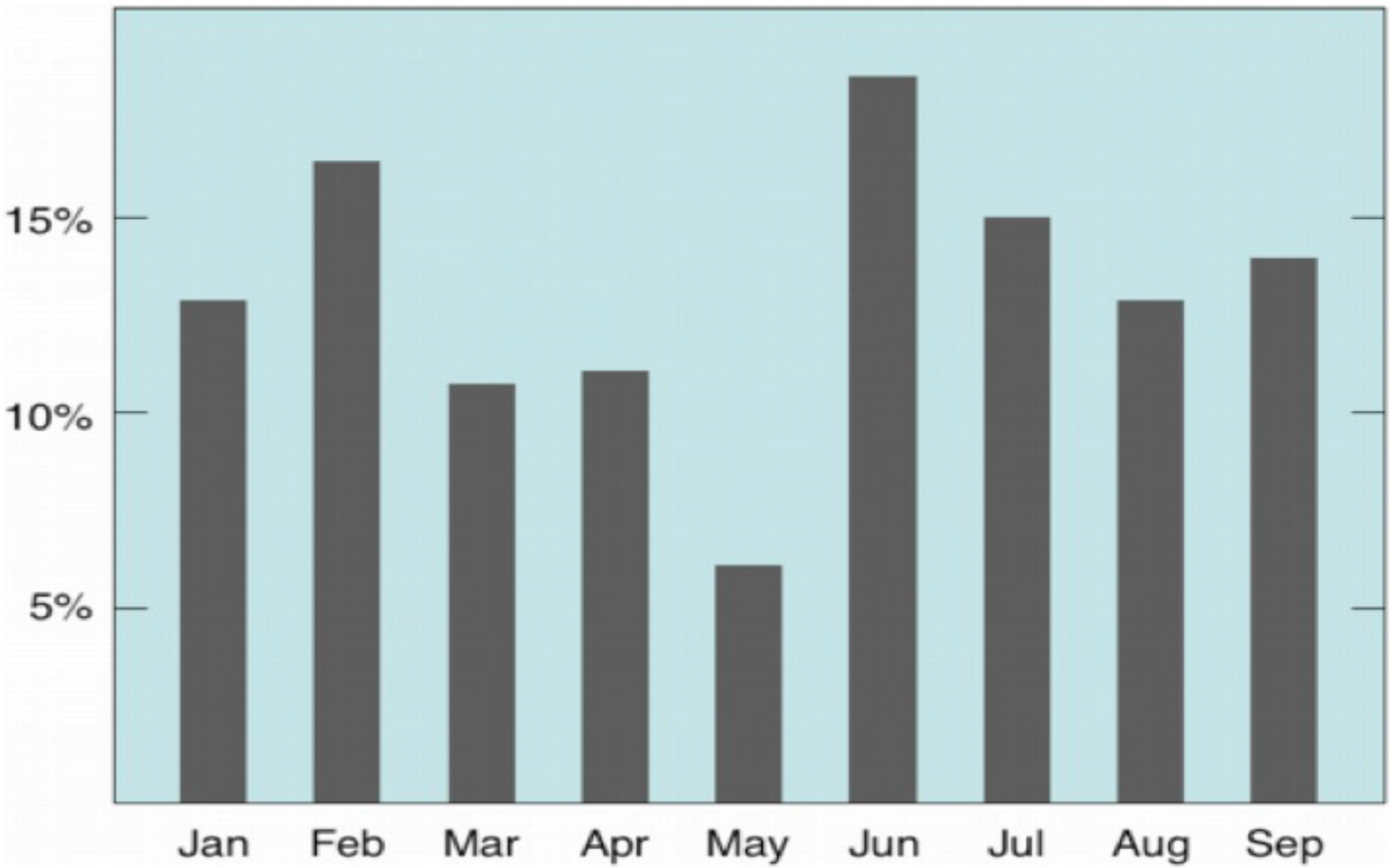
MAXIMIZE THE RATIO OF

$$\frac{(\textbf{INK}\ \text{USED TO SHOW DATA})}{(\textbf{TOTAL INK}\ \text{USED})}$$

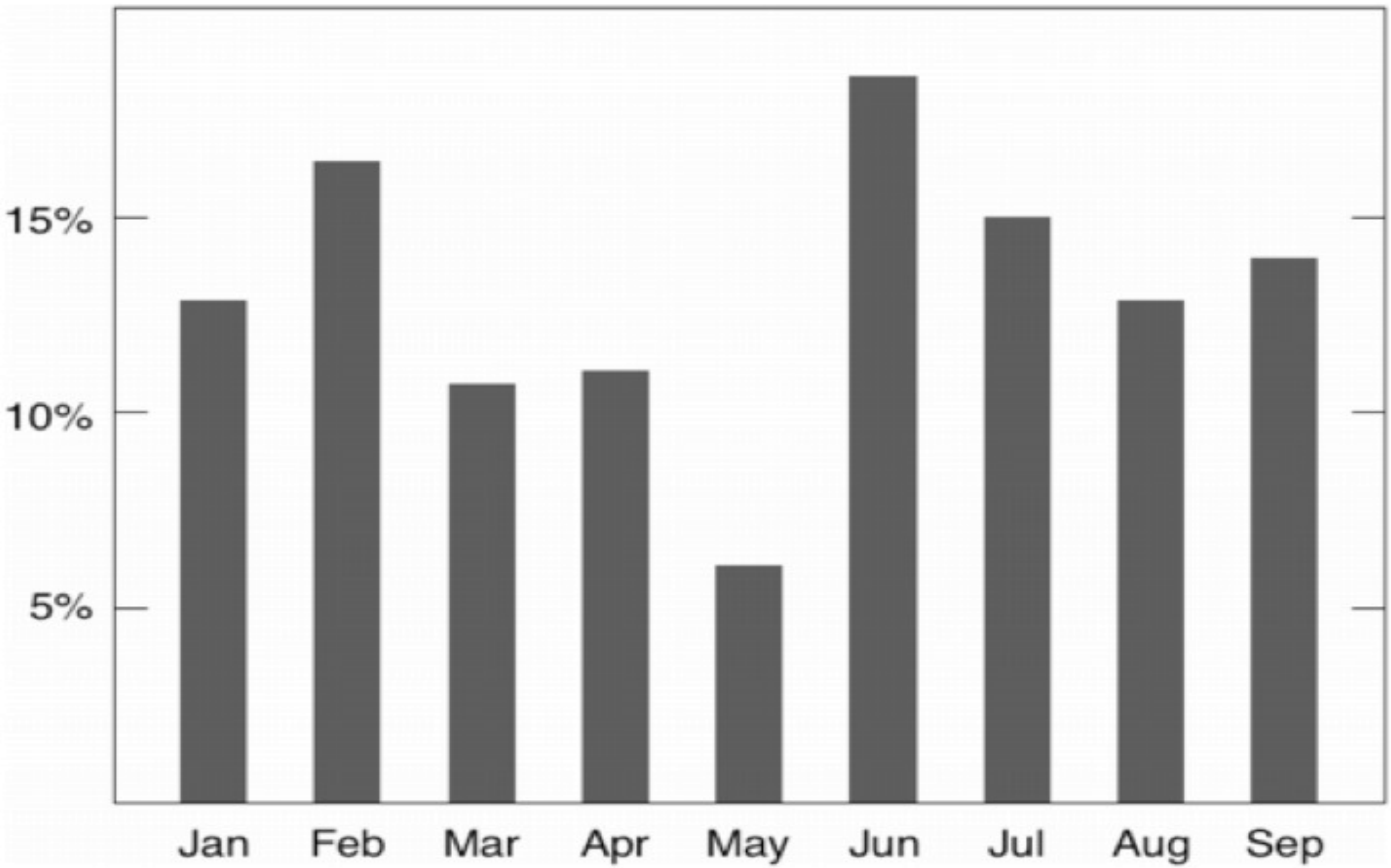
DATA-INK RATIO



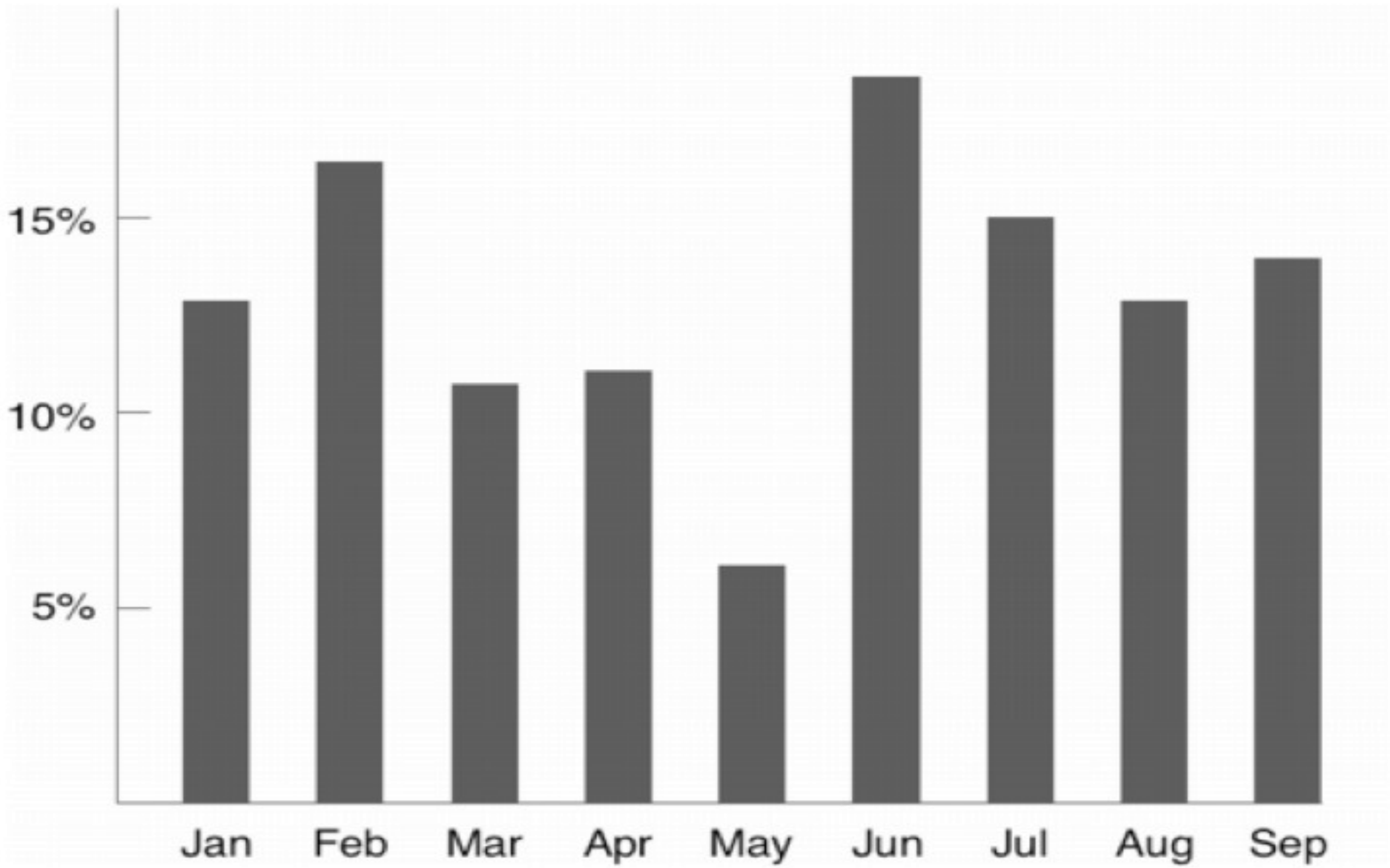
DATA-INK RATIO



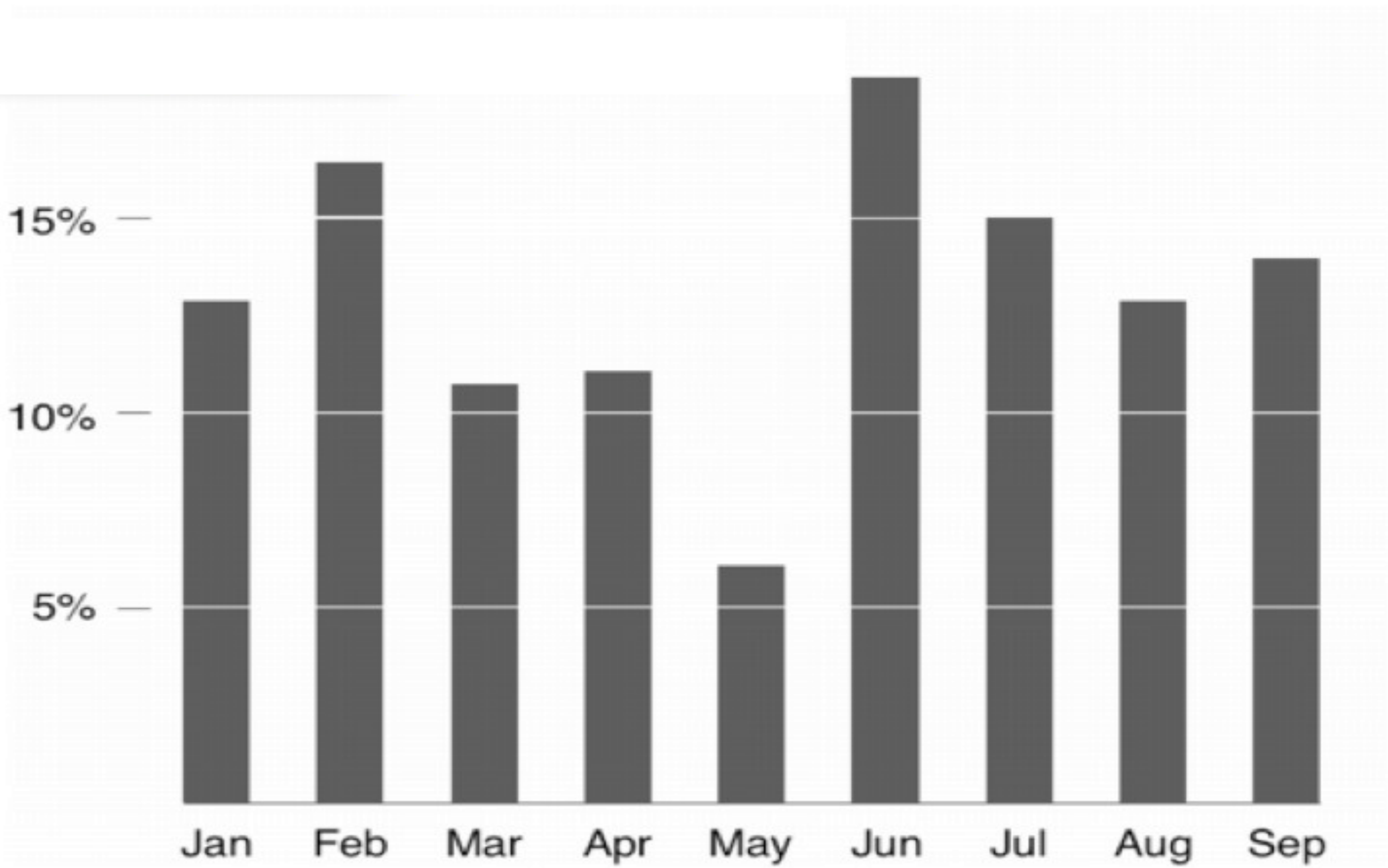
DATA-INK RATIO



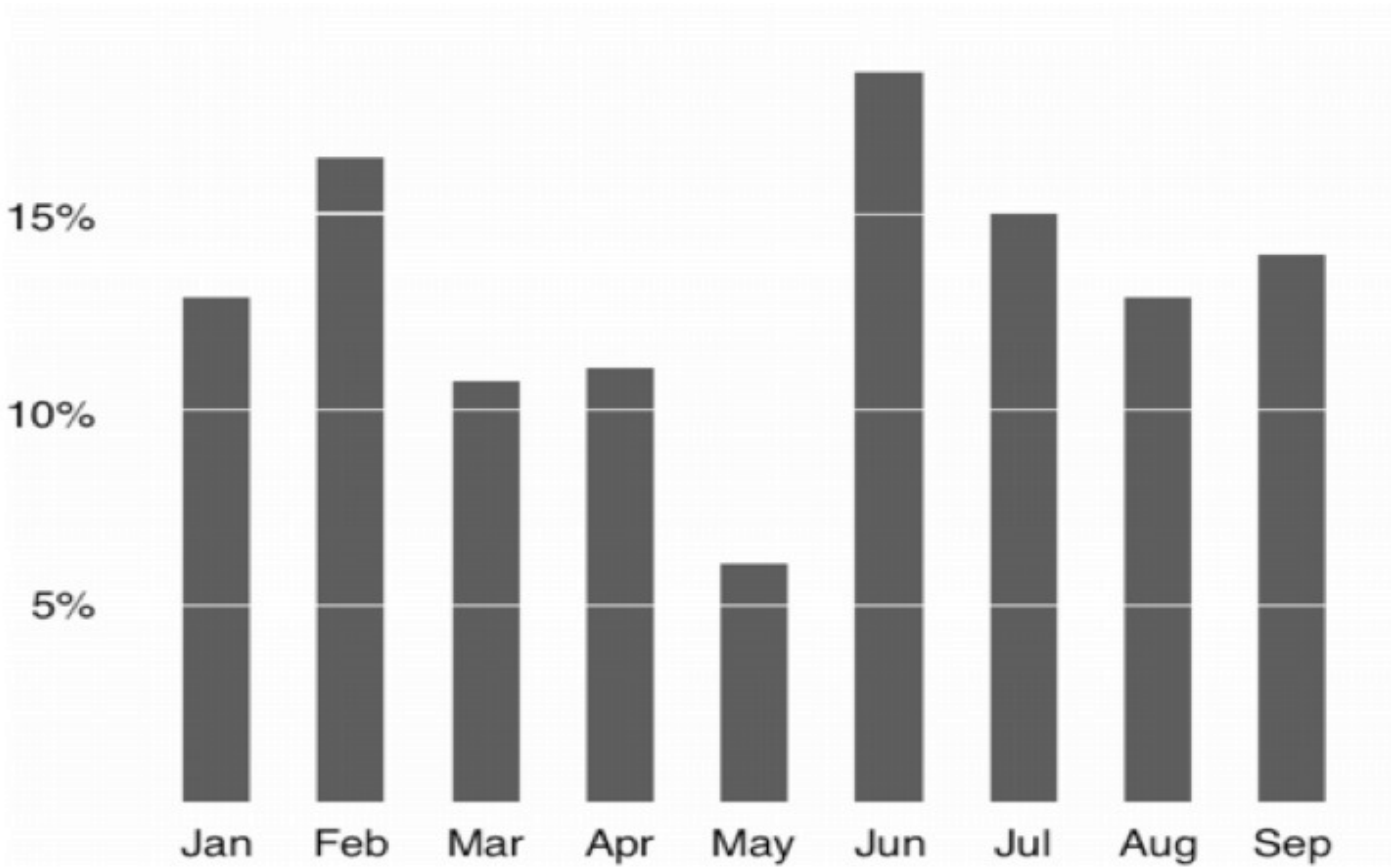
DATA-INK RATIO



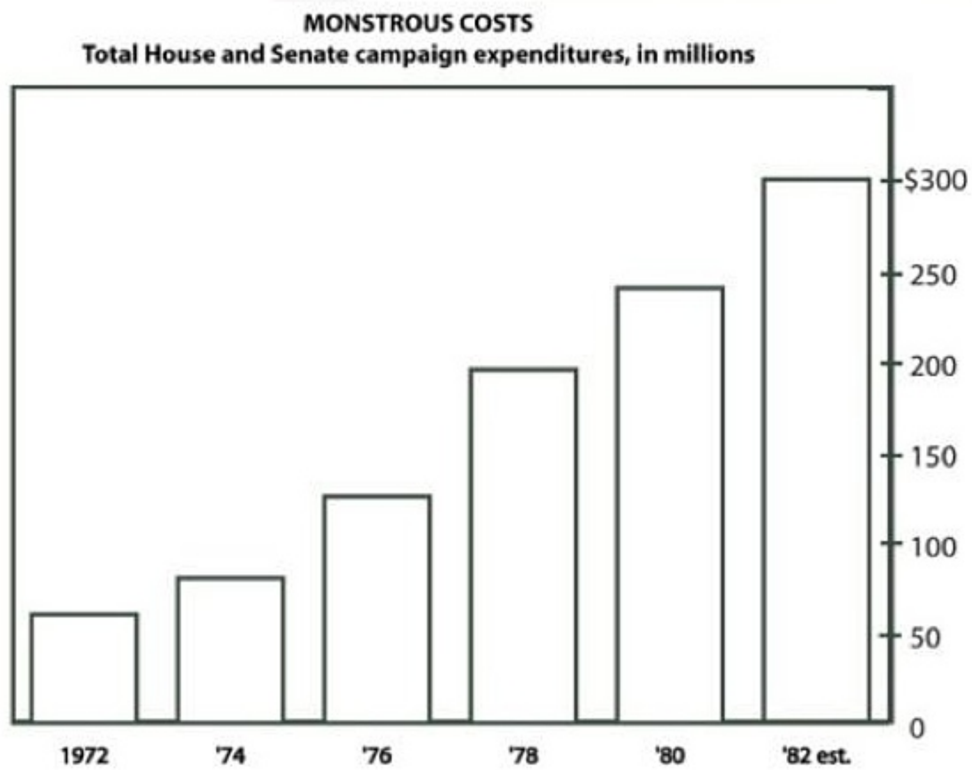
DATA-INK RATIO

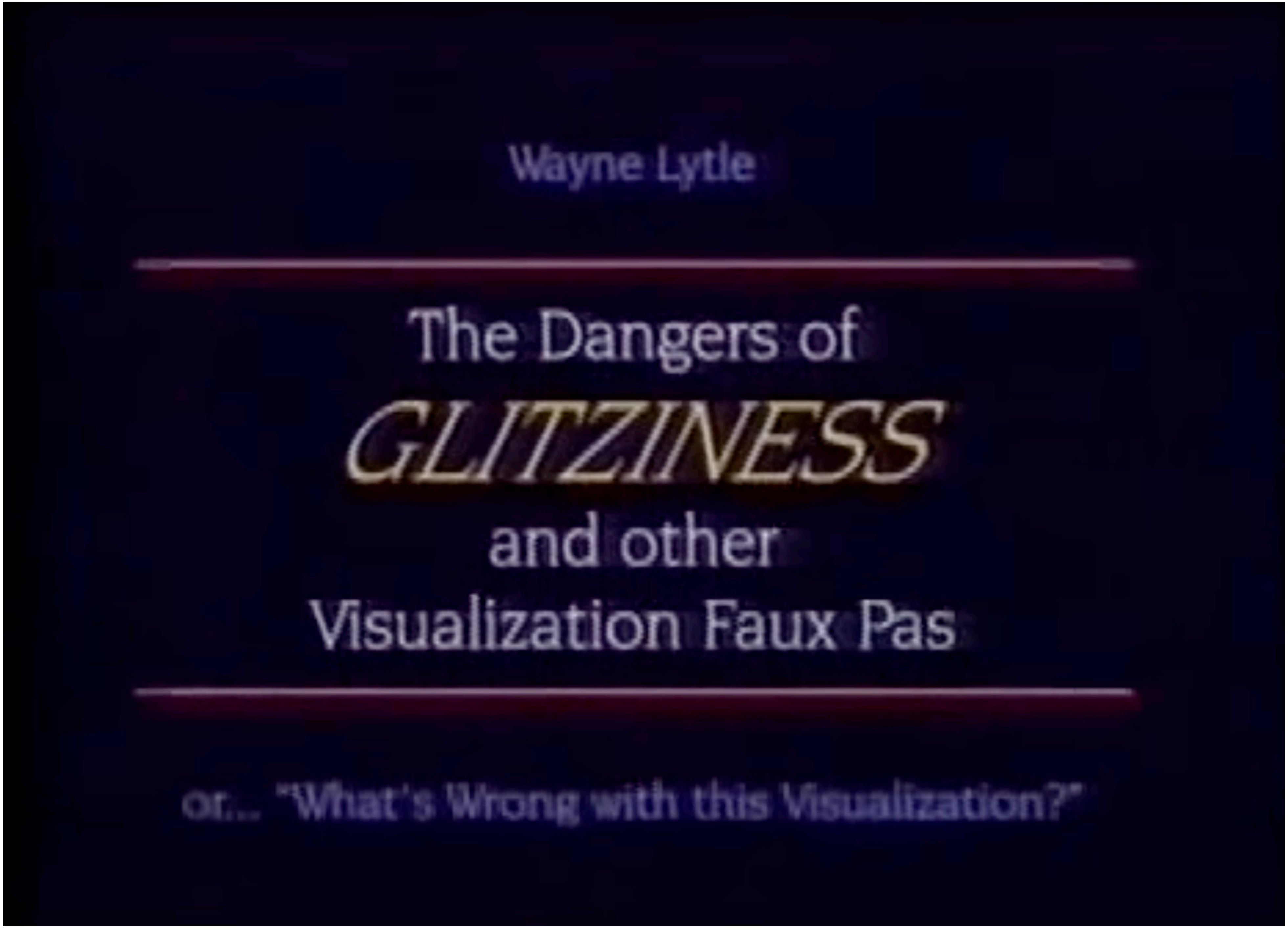


DATA-INK RATIO



MINIMIZE CHART JUNK



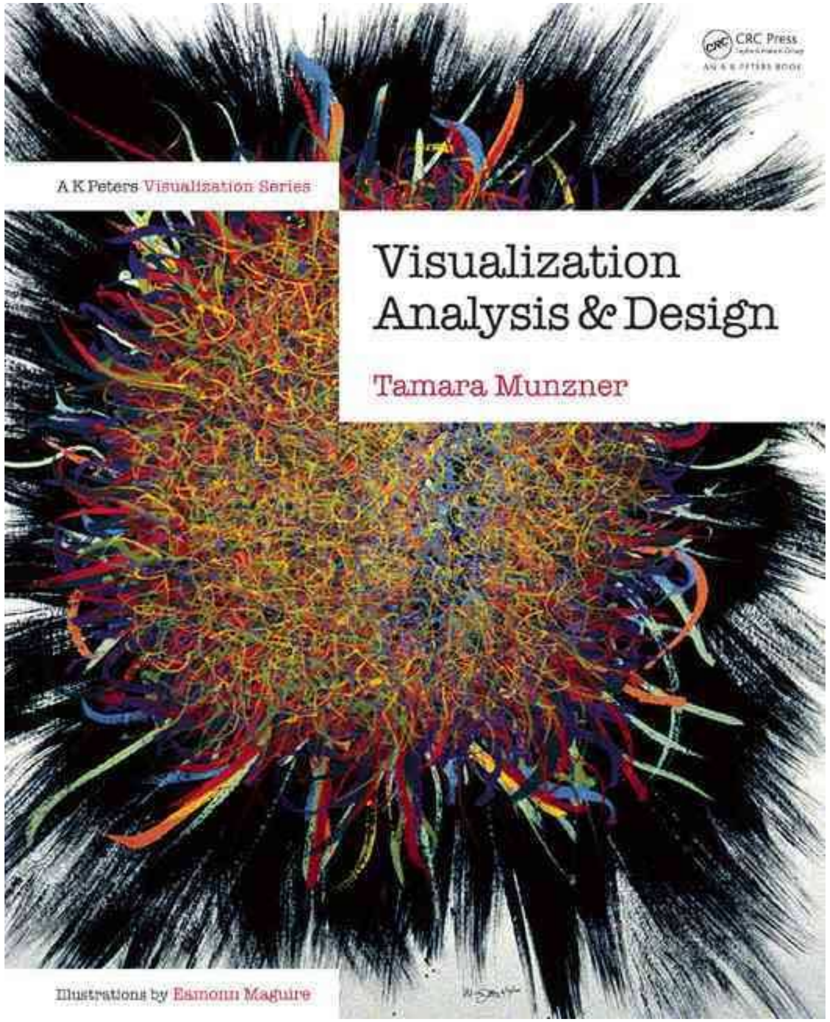

Wayne Lytle
The Dangers of
GLITZINESS
and other
Visualization Faux Pas
or... "What's Wrong with this Visualization?"

TUFTE'S INTEGRITY PRINCIPLES

EDWARD TUFTE

- MAXIMIZE THE DATA-INK RATIO
- AVOID CHART JUNK (*SOMETIMES*)
- LAYER INFORMATION
- MAXIMIZE THE DATA DENSITY
  - *SHRINK THE GRAPHICS*
  - *MAXIMIZE THE AMOUNT OF DATA SHOWN (SOMETIMES)*

READINGS



CRC Press
A K Peters Visualization Series
Visualization
Analysis & Design
Tamara Munzner
Illustrations by Eamonn Maguire

ACKNOWLEDGEMENTS

Slides in were inspired and adapted from slides by

- Nicolai Marquardt (University College London)
- Uta Hinrichs (University of St. Andrews)
- Saul Greenberg (University of Calgary)