

CS 519: Scientific Visualization

Information Visualization: Tables

Eric Shaffer

Some (many!) slides adapted from work by Professor Tamara Munzner at the University of British Columbia

Tables

- The most ubiquitous types of InfoVis data ?
- $T = \{r_i\}_{i=1..m}$, $r = \{c_j\}_{j=1..n}$, $c_j \in D_j$, $D_j \subseteq \text{Numerical} \cup \text{Ordinal} \cup \text{Categorical}$
 - columns can have different types
 - rows are not uniquely ordered

Example: stock exchange data

id	category	name	date	time	open	high	low	close
636	sif	SIF1	2004-11-29	13:00	0.800000	0.800000	0.800000	0.800000
635	sif	SIF1	2004-11-29	14:00	0.800000	0.800000	0.800000	0.800000
633	sif	SIF1	2004-11-29	16:00	0.795000	0.795000	0.795000	0.795000
630	sif	SIF1	2004-11-30	14:00	0.795000	0.795000	0.795000	0.795000
632	sif	SIF1	2004-11-30	12:00	0.800000	0.800000	0.795000	0.795000
631	sif	SIF1	2004-11-30	13:00	0.795000	0.795000	0.795000	0.795000
628	sif	SIF1	2004-11-30	16:00	0.795000	0.795000	0.795000	0.795000
629	sif	SIF1	2004-11-30	15:00	0.795000	0.795000	0.795000	0.795000
627	sif	SIF1	2005-00-02	12:00	0.785000	0.790000	0.785000	0.790000
626	sif	SIF1	2005-00-02	13:00	0.790000	0.795000	0.790000	0.795000
625	sif	SIF1	2005-00-02	14:00	0.795000	0.795000	0.795000	0.795000
624	sif	SIF1	2005-00-02	15:00	0.800000	0.800000	0.800000	0.800000
620	sif	SIF1	2005-00-03	15:00	0.795000	0.795000	0.795000	0.795000
623	sif	SIF1	2005-00-03	12:00	0.795000	0.795000	0.795000	0.795000
622	sif	SIF1	2005-00-03	13:00	0.795000	0.795000	0.795000	0.795000
621	sif	SIF1	2005-00-03	14:00	0.795000	0.795000	0.795000	0.795000
619	sif	SIF1	2005-00-03	16:00	0.795000	0.795000	0.795000	0.795000
618	sif	SIF1	2005-00-06	11:00	0.790000	0.790000	0.790000	0.790000
614	sif	SIF1	2005-00-06	15:00	0.795000	0.795000	0.795000	0.795000
617	sif	SIF1	2005-00-06	12:00	0.795000	0.795000	0.795000	0.795000
616	sif	SIF1	2005-00-06	13:00	0.795000	0.795000	0.795000	0.795000
615	sif	SIF1	2005-00-06	14:00	0.795000	0.795000	0.795000	0.795000
613	sif	SIF1	2005-00-06	16:00	0.795000	0.795000	0.795000	0.795000
609	sif	SIF1	2005-00-07	14:00	0.790000	0.795000	0.790000	0.795000
612	sif	SIF1	2005-00-07	11:00	0.795000	0.795000	0.795000	0.795000
611	sif	SIF1	2005-00-07	12:00	0.795000	0.795000	0.795000	0.795000
610	sif	SIF1	2005-00-07	13:00	0.790000	0.790000	0.790000	0.790000
608	sif	SIF1	2005-00-07	15:00	0.790000	0.790000	0.790000	0.790000
606	sif	SIF1	2005-00-08	13:00	0.795000	0.795000	0.795000	0.795000
607	sif	SIF1	2005-00-08	12:00	0.790000	0.790000	0.790000	0.790000
605	sif	SIF1	2005-00-08	14:00	0.795000	0.795000	0.795000	0.795000

Design Choices For Tabular Visualization

④ Express Values



⑤ Separate, Order, Align Regions

→ Separate



→ Order

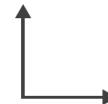


→ Align



⑥ Axis Orientation

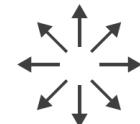
→ Rectilinear



→ Parallel

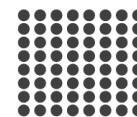


→ Radial



⑦ Layout Density

→ Dense



→ Space-Filling



→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume



→ Many Keys
Recursive Subdivision

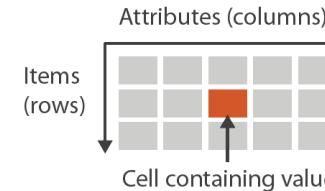


Keys and Values

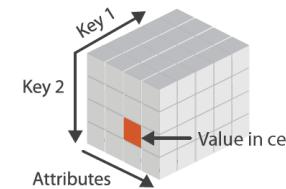
Keys and values

- key
 - independent attribute
 - used as unique index to look up items
 - simple tables: 1 key
 - multidimensional tables: multiple keys
- value
 - dependent attribute, value of cell
- classify arrangements by key count
 - 0, 1, 2, many...

→ Tables



→ Multidimensional Table



⇒ Express Values



→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume



→ Many Keys
Recursive Subdivision

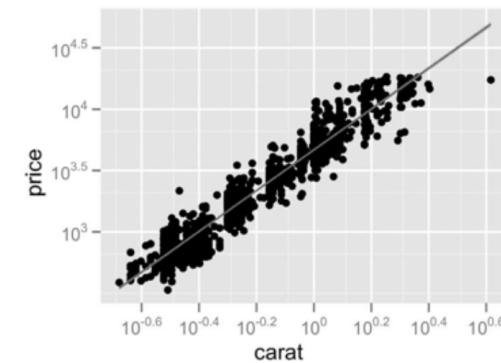
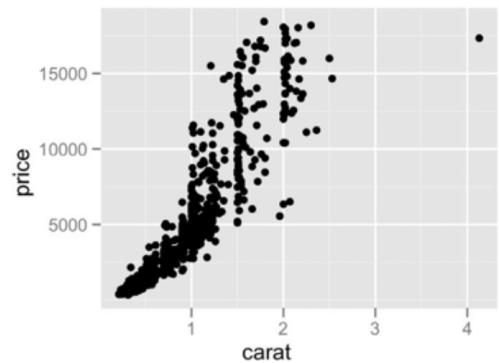
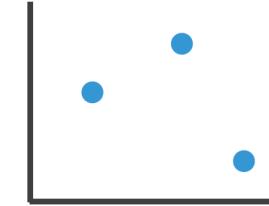


Idiom: Scatterplot

Idiom: **scatterplot**

- **express** values
 - quantitative attributes
- no keys, only values
 - data
 - 2 quant attrs
 - mark: points
 - channels
 - horiz + vert position
 - tasks
 - find trends, outliers, distribution, correlation, clusters
 - scalability
 - hundreds of items

→ Express Values



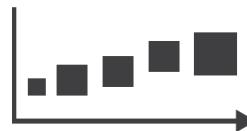
[*A layered grammar of graphics*. Wickham. *Journ. Computational and Graphical Statistics* 19:1 (2010), 3–28.]

Some Keys: Categorical Regions

→ Separate



→ Order



→ Align



- **regions:** contiguous bounded areas distinct from each other
 - using space to *separate* (proximity)
 - following expressiveness principle for categorical attributes
- use ordered attribute to *order* and *align* regions

→ 1 Key
List



→ 2 Keys
Matrix



→ 3 Keys
Volume



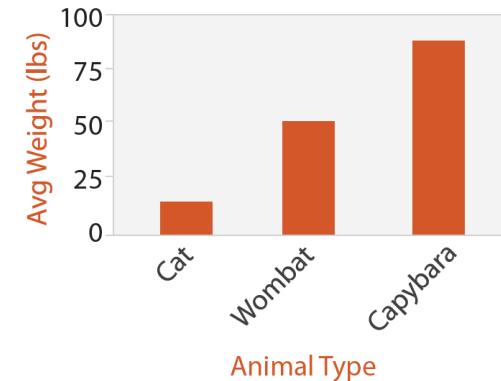
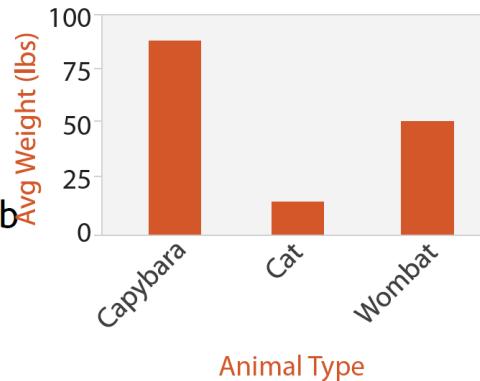
→ Many Keys
Recursive Subdivision



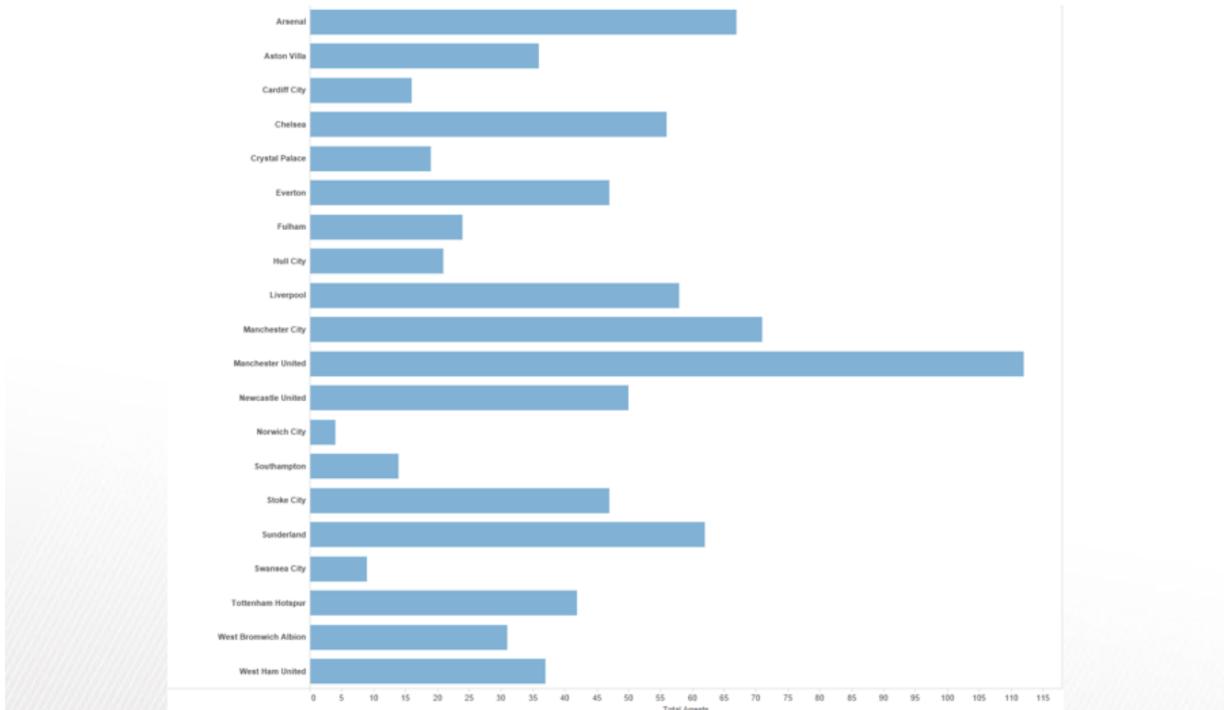
Idiom: Bar Charts

Idiom: **bar chart**

- one key, one value
 - data
 - 1 categ attrib, 1 quant attrib
 - mark: lines
 - channels
 - length to express quant value
 - spatial regions: one per mark
 - separated horizontally, aligned vertically
 - ordered by quant attrib
 - » by label (alphabetical), by length attrib (data-driven)
 - task
 - compare, lookup values
 - scalability
 - dozens to hundreds of levels for key attrib



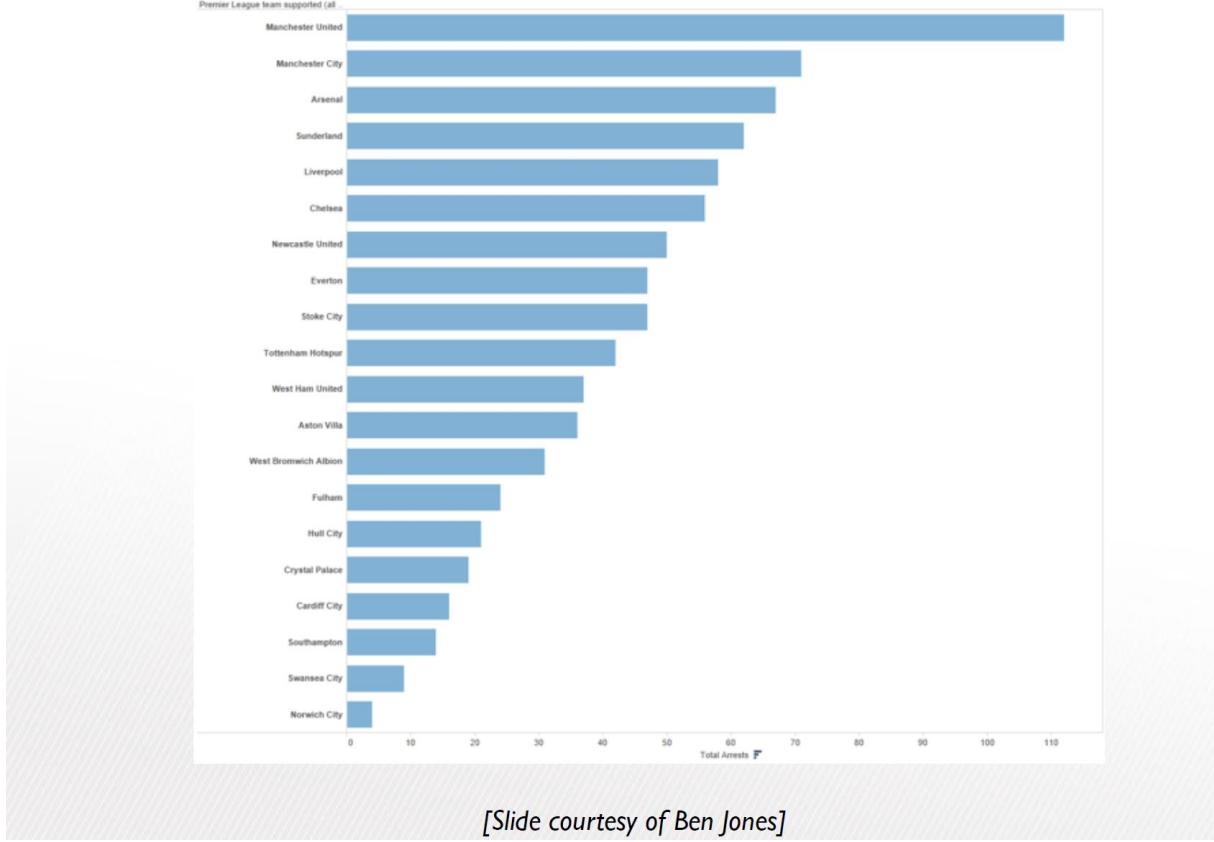
Separated and Aligned but not Ordered



LIMITATION: Hard to know rank. What's the 4th most? The 7th?

[Slide courtesy of Ben Jones]

Separated and Aligned and Ordered



Separated but not Ordered nor Aligned



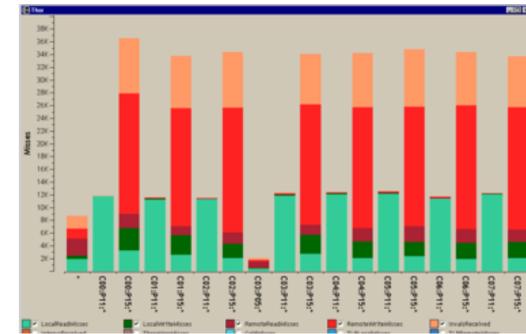
LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]

Idiom: Stacked Bar Chart

Idiom: **stacked bar chart**

- one more key
 - data
 - 2 categ attrib, 1 quant attrib
 - mark: vertical stack of line marks
 - **glyph**: composite object, internal structure from multiple marks
 - channels
 - length and color hue
 - spatial regions: one per glyph
 - aligned: full glyph, lowest bar component
 - unaligned: other bar components
 - task
 - part-to-whole relationship
 - scalability
 - several to one dozen levels for stacked attrib

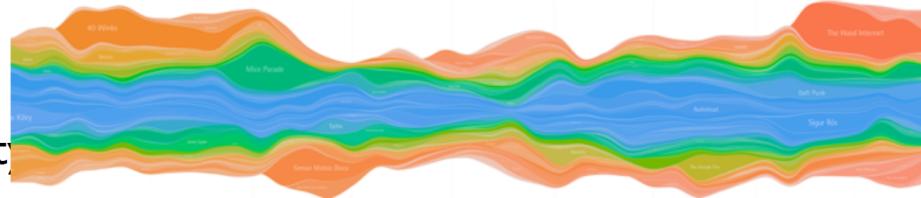


[Using Visualization to Understand the Behavior of Computer Systems. Bosch. Ph.D. thesis, Stanford Computer Science, 2001.]

Idiom: StreamGraph (aka ThemeRiver)

Idiom: **streamgraph**

- generalized stacked graph
 - emphasizing horizontal continuity
 - vs vertical items
 - data
 - 1 categ key attrib (artist)
 - 1 ordered key attrib (time)
 - 1 quant value attrib (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

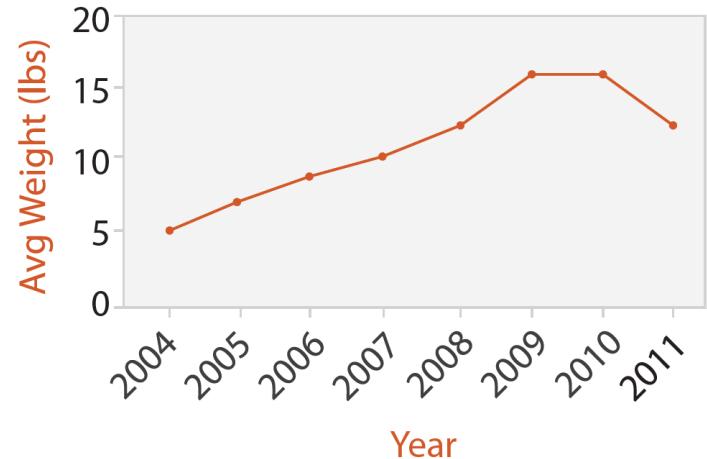


[*Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).*]

Idiom: Line Chart

Idiom: **line chart / dot plot**

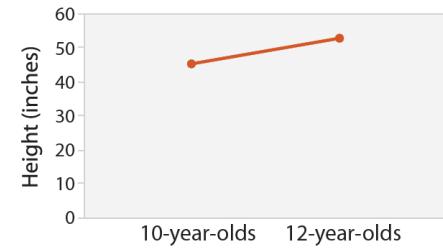
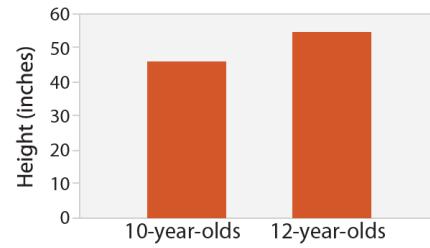
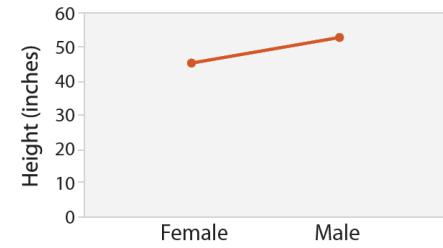
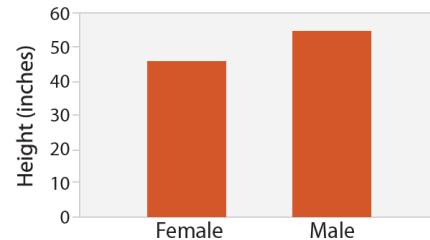
- one key, one value
 - data
 - 2 quant attrs
 - mark: points
 - line connection marks between them
 - channels
 - aligned lengths to express quant value
 - separated and ordered by key attrib into horizontal regions
 - task
 - find trend
 - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
 - scalability
 - hundreds of key levels, hundreds of value levels



Bar vs Line

Choosing bar vs line charts

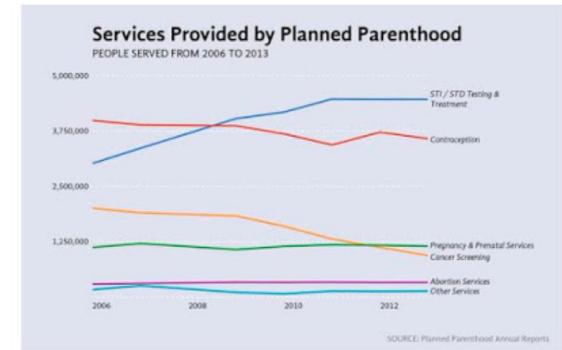
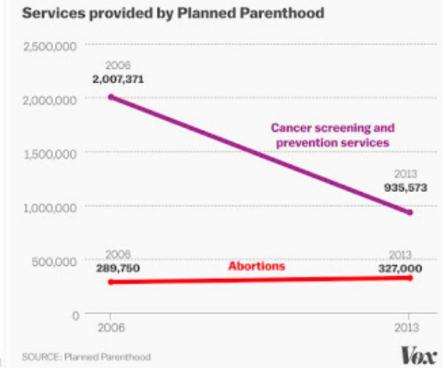
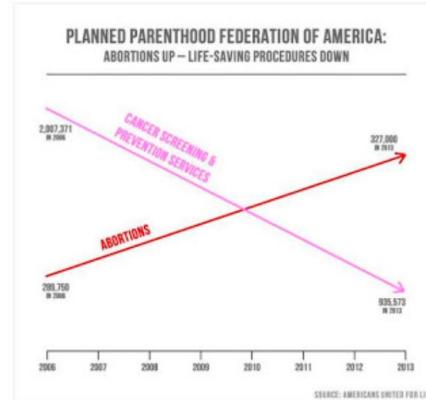
- depends on type of key attrib
 - bar charts if categorical
 - line charts if ordered
- do not use line charts for categorical key attrs
 - violates expressiveness principle
 - implication of trend so strong that it overrides semantics!
 - “The more male a person is, the taller he/she is”



after [Bars and Lines: A Study of Graphic Communication.
Zacks and Tversky. Memory and Cognition 27:6 (1999),
1073–1079.]

Chart Axes

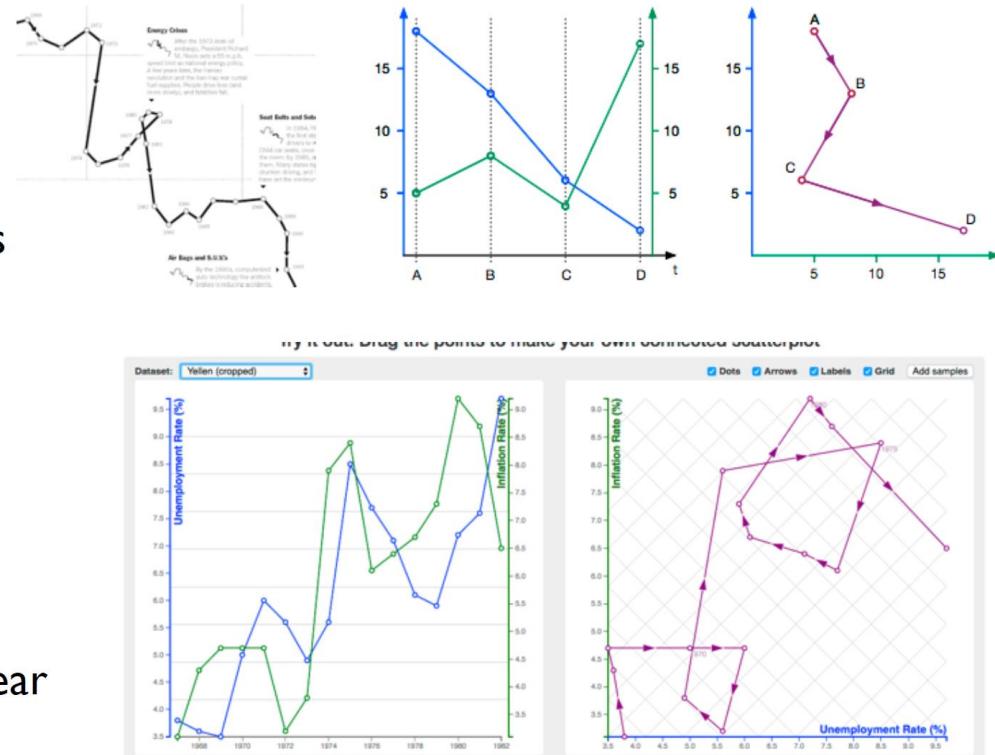
- labelled axis is critical
- avoid cropping y-axis
 - include 0 at bottom left
 - or slope misleads
- dual axes controversial
 - acceptable if commensurate
 - beware, very easy to mislead!



Idiom: Connected Scatterplots

Idiom: **connected scatterplots**

- scatterplot with line connection marks
 - popular in journalism
 - horiz + vert axes: value attrs
 - line connection marks: temporal order
 - alternative to dual-axis charts
 - horiz: time
 - vert: two value attrs
- empirical study
 - engaging, but correlation unclear



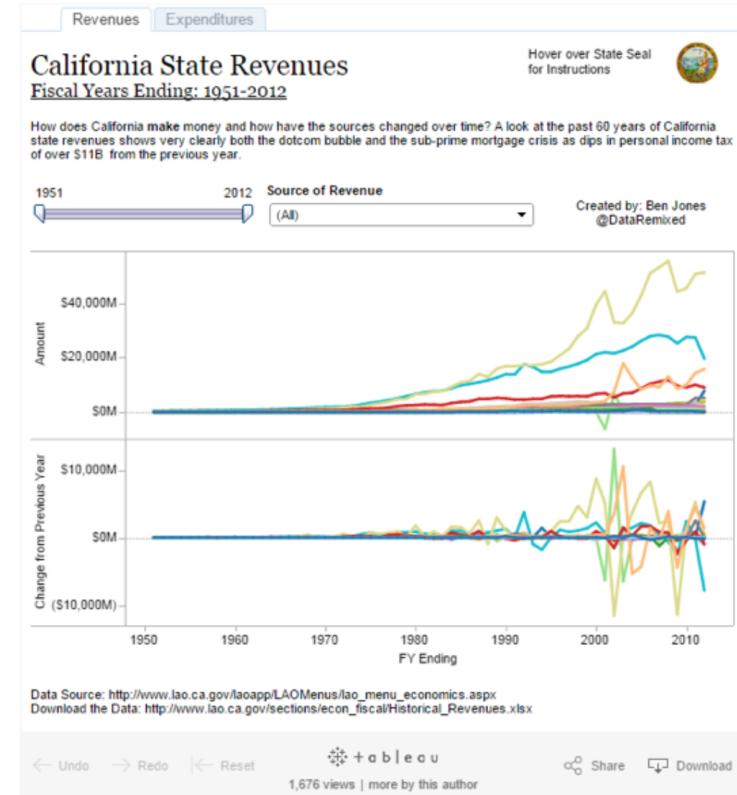
[The Connected Scatterplot for Presenting Paired Time Series.
Haroz, Kosara and Franconeri. IEEE TVCG 22(9):2174-86, 2016.]

http://steveharoz.com/research/connected_scatterplot/

Idiom: Indexed Line Charts

Idiom: **Indexed line charts**

- data: 2 quant attires
 - 1 key + 1 value
- derived data: new quant value attrib
 - index
 - plot instead of original value
- task: show change over time
 - principle: normalized, not absolute
- scalability
 - same as standard line chart

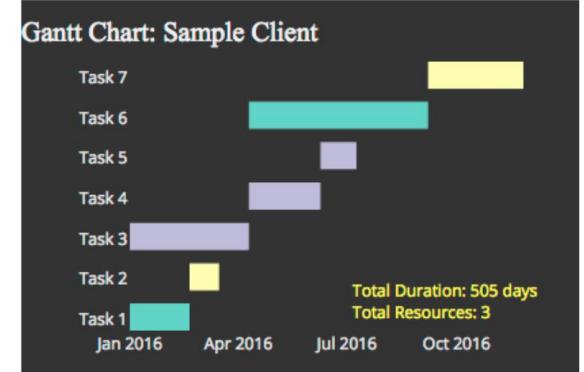


<https://public.tableau.com/profile/ben.jones#!/vizhome/CAStrateRevenues/Revenues> 22

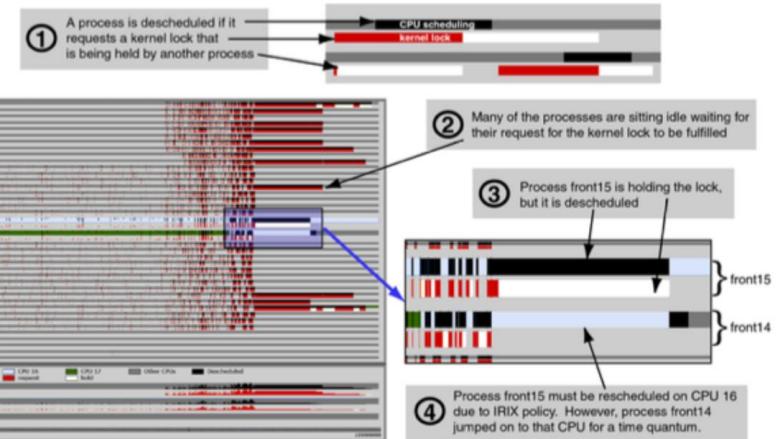
Idiom: Gantt Charts

Idiom: **Gantt charts**

- one key, two (related) values
 - data
 - 1 categ attrib, 2 quant attribs
 - mark: line
 - length: duration
 - channels
 - horiz position: start /end times
 - horiz length: duration
 - task
 - emphasize temporal overlaps, start/end dependencies between items
 - scalability
 - dozens of key levels
 - hundreds of value levels
 - hundreds of value levels



<https://www.r-bloggers.com/gantt-charts-in-r-using-plotly/>



[Performance Analysis and Visualization of Parallel Systems Using SimOS and Rivet: A Case Study. Bosch, Stolte, Stoll, Rosenblum, and Hanrahan. Proc. HPCA 2000.]

23

[Performance Analysis and Visualization of Parallel Systems Using SimOS and Rivet: A Case Study. Bosch, Stolte, Stoll, Rosenblum, and Hanrahan. Proc. HPCA 2000.]

23

Idiom: Heat Map

Idiom: **heatmap**

- two keys, one value

- data

- 2 categ attrs (gene, experimental condition)
 - 1 quant attrib (expression levels)

- marks: area

- separate and align in 2D matrix
 - indexed by 2 categorical attributes

- channels

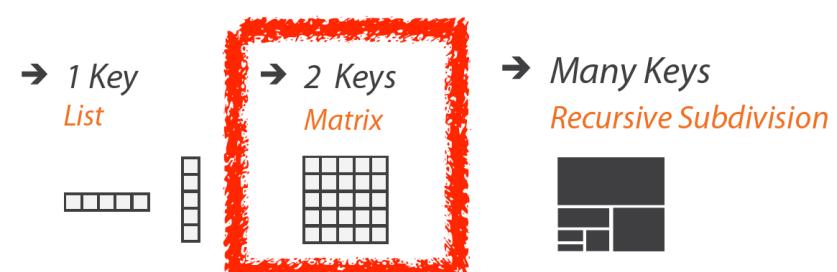
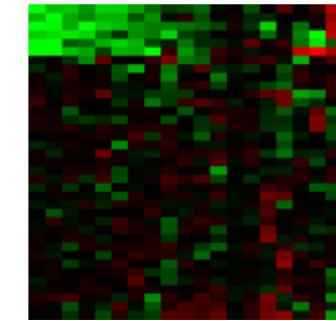
- color by quant attrib
 - (ordered diverging colormap)

- task

- find clusters, outliers

- scalability

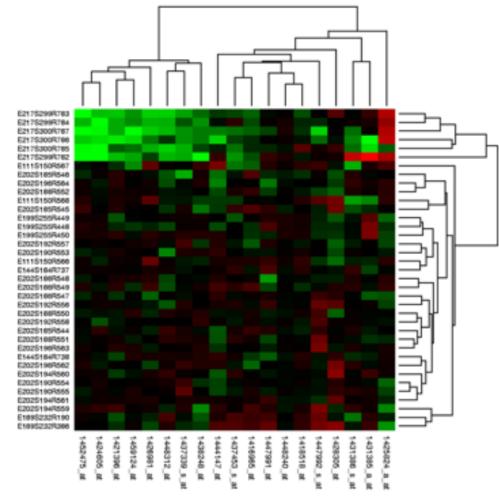
- 1M items, 100s of categ levels, ~10 quant attrib levels



Idiom: Cluster Heat Map

Idiom: **cluster heatmap**

- in addition
 - derived data
 - 2 cluster hierarchies
 - dendrogram
 - parent-child relationships in tree with connection line marks
 - leaves aligned so interior branch heights easy to compare
 - heatmap
 - marks (re-)ordered by cluster hierarchy traversal

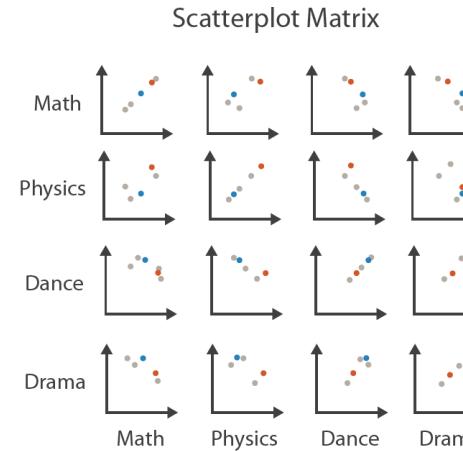


Idiom: Scatter Plot Matrix

Idioms: **scatterplot matrix, parallel coordinates**

- scatterplot matrix (SPLOM)

- rectilinear axes, point mark
- all possible pairs of axes
- scalability
 - one dozen attrs
 - dozens to hundreds of items



- parallel coordinates

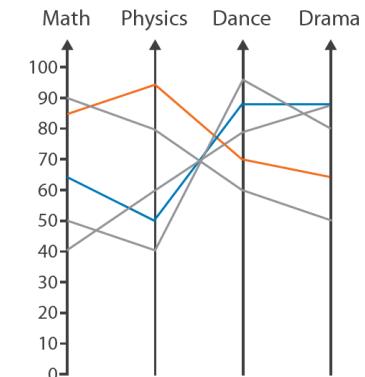
- parallel axes, jagged line representing item
- rectilinear axes, item as point

- axis ordering is major challenge

- scalability

- dozens of attrs
- hundreds of items

Parallel Coordinates



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	

after [Visualization Course Figures. McGuffin, 2014. <http://www.michaelmcguffin.com/courses/vis/>]

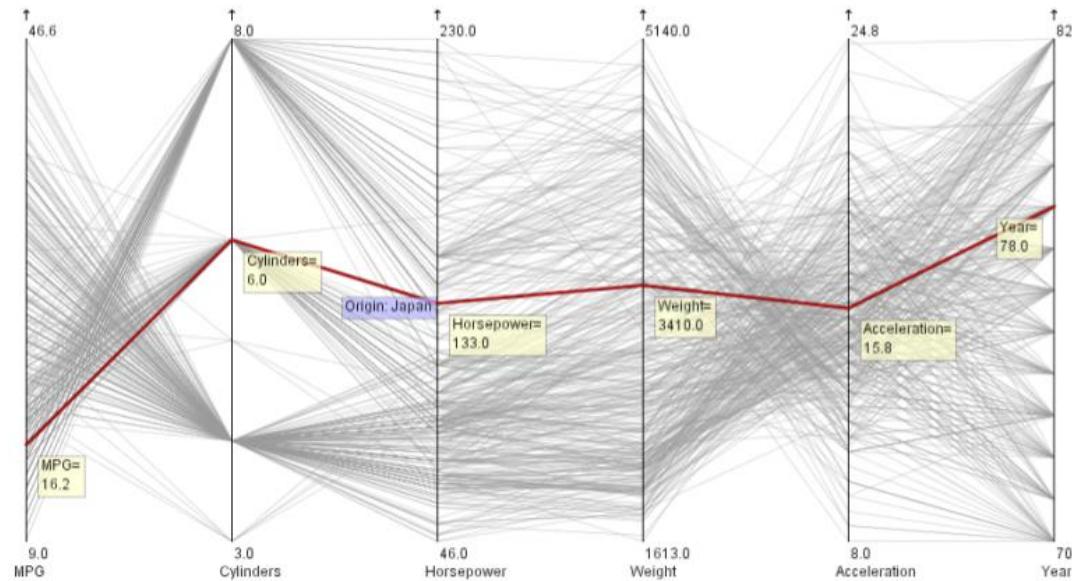
Tables: Parallel Coordinates

Take a table

- rows: car brands
- columns: car parameters (MPG, cylinders, horsepower, weight, acceleration, fabrication year)

Parallel coordinates [Inselberg and Dinsdale '90]

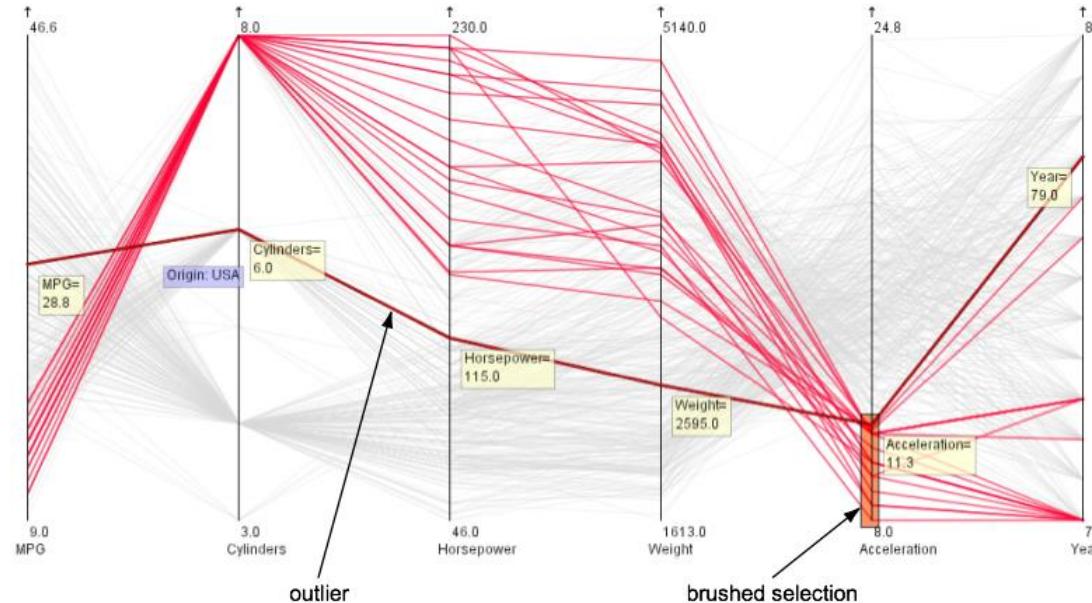
- columns: different y axes
- cells: points on their corresponding axes
- rows: polylines connecting their points
- correlations: ‘bundles’ of close lines



Parallel Coordinates

Selection

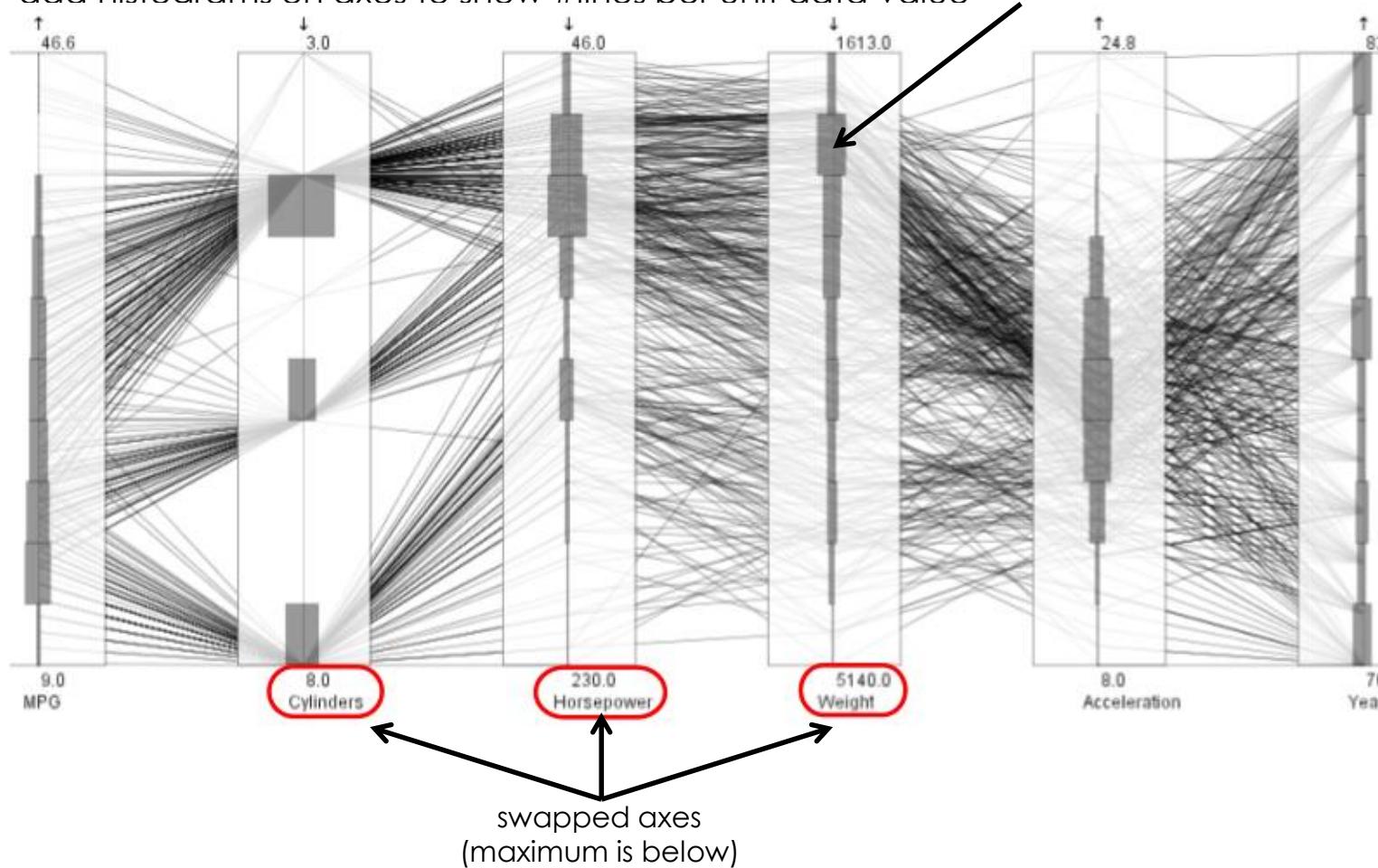
- use mouse to select attribute ranges on axes
- highlight all rows (lines) passing through selection
- supports queries such as
 - show all cars with a low acceleration
 - find what attributes (e.g. MPG, cylinders, weight, ...) low-acceleration cars have



Parallel Coordinates

Enhancements

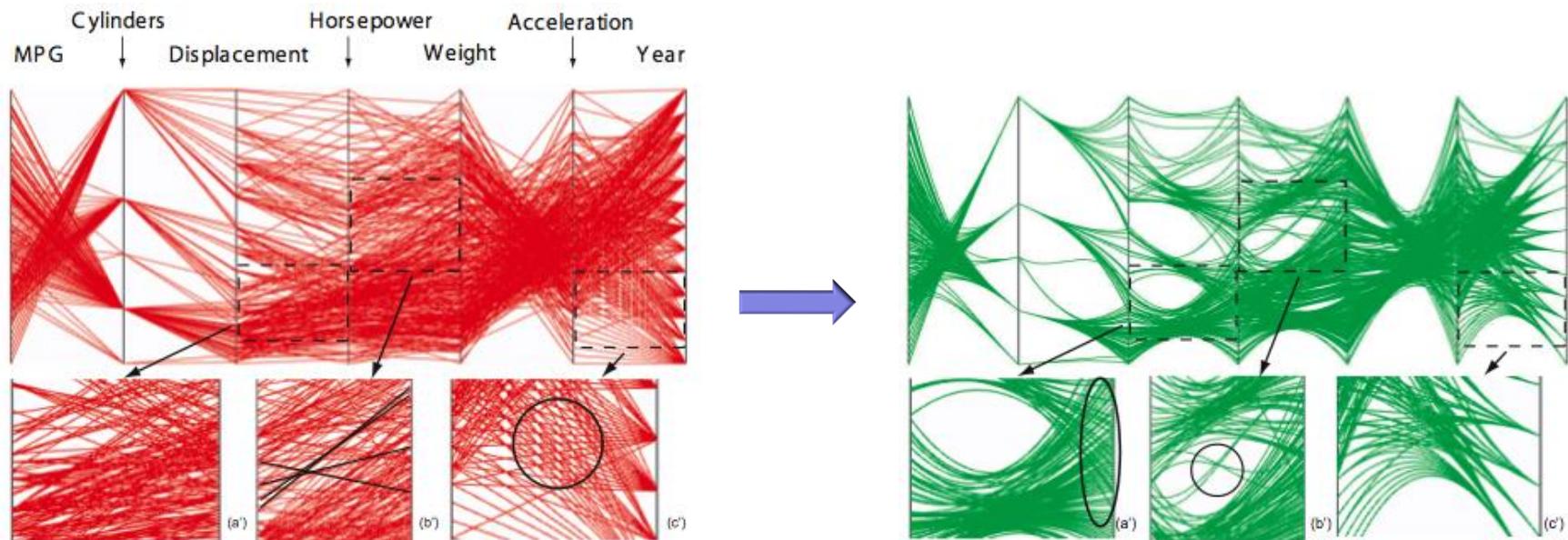
- permute axes (horizontally) to and swap their direction (vertically) minimize crossings
- add histoarams on axes to show #lines per unit data value histograms



Parallel Coordinates

Clutter reduction [Zhou et al. '08]

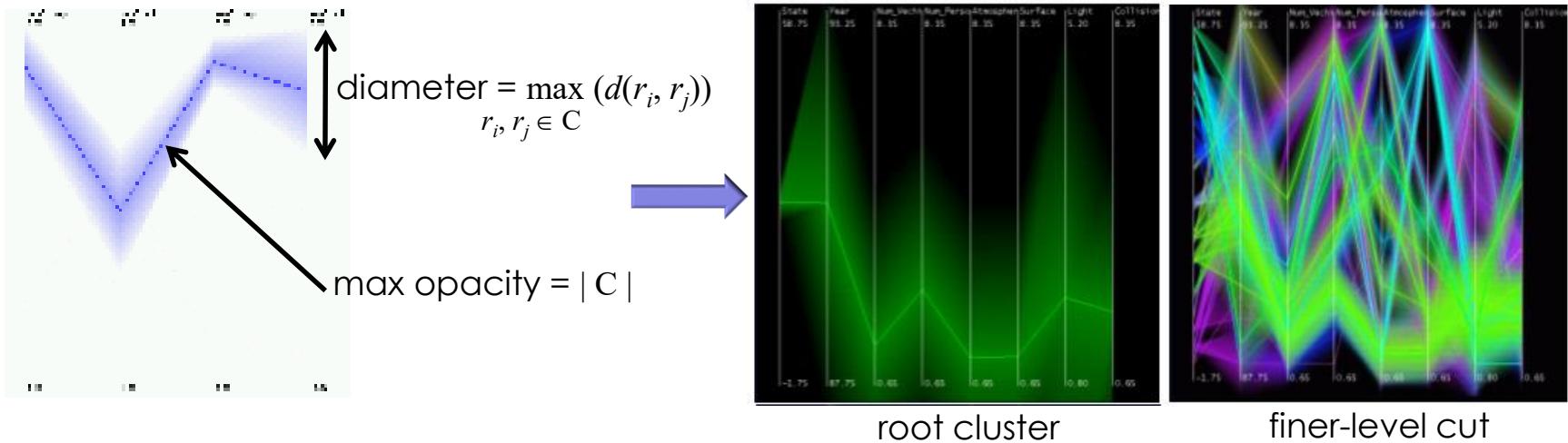
- define an energy function
- energy is lowest when
 - line curvature is low
 - relative ordering of curves is preserved
 - a curve is close to its neighbor curves



Parallel Coordinates

Hierarchical parallel coordinates [Fua et al. '99]

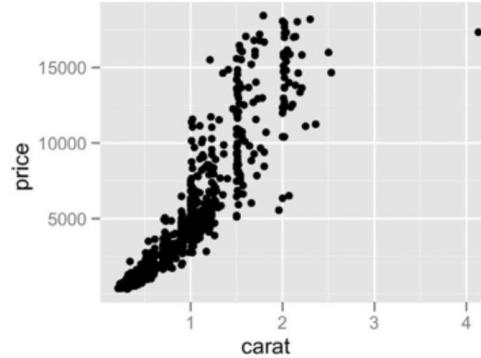
- reduce clutter for very large datasets ($10^6..10^9$ rows)
- hierarchically cluster rows r_i
 - create a cluster $C_i = \{r_i\}$ for each row, $S = \{C_i\}$
 - find two most similar clusters C_i, C_j using an Euclidean distance metric $d(r_i, r_j) = \sum_k (r_{ik} - r_{jk})^2$
 - build parent cluster $C = (C_i, C_j)$, $S = S \setminus (C_i \cup C_j) \cup C$
 - repeat from step 2 until $S = \{\text{root cluster}\}$
- select a ‘cut’ K in the cluster tree at desired detail level
- visualize each cluster $C \in K$ with an opacity band which encodes cluster size and diameter



Finding Correlations

Task: Correlation

- scatterplot matrix
 - positive correlation
 - diagonal low-to-high
 - negative correlation
 - diagonal high-to-low
 - uncorrelated
- parallel coordinates
 - positive correlation
 - parallel line segments
 - negative correlation
 - all segments cross at halfway point
 - uncorrelated
 - scattered crossings



[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

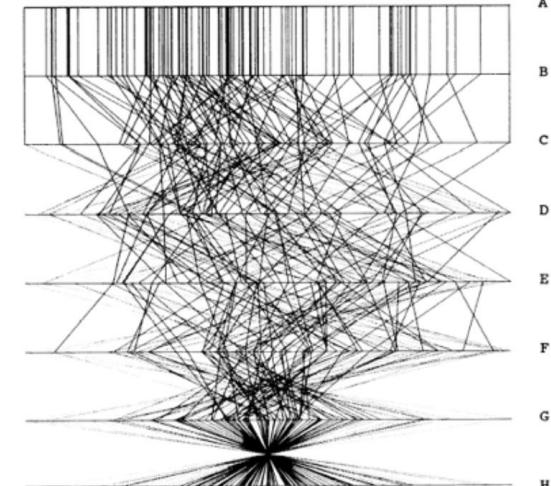
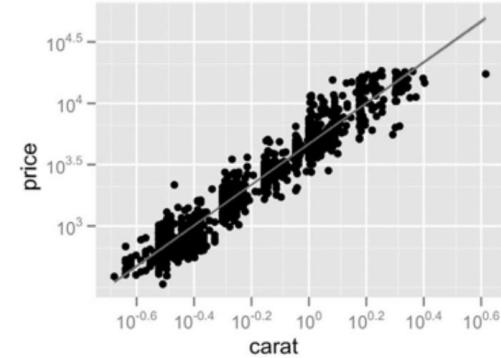
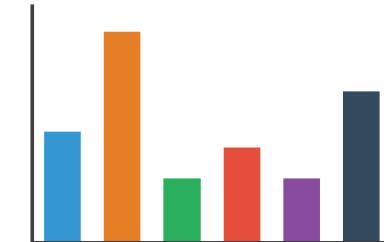
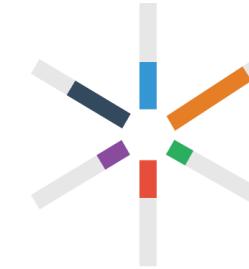


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8$, and -1 .

Idiom: Radial Bar Chart

Idioms: **radial bar chart, star plot**

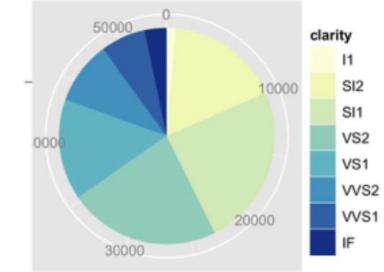
- radial bar chart
 - radial axes meet at central ring, line mark
- star plot
 - radial axes, meet at central point, line mark
- bar chart
 - rectilinear axes, aligned vertically
- accuracy
 - length unaligned with radial
 - less accurate than aligned with rectilinear



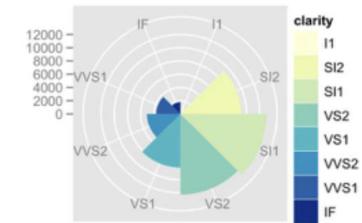
Idiom: Pie Chart

Idioms: pie chart, polar area chart

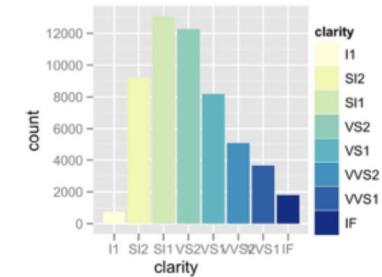
- pie chart
 - area marks with angle channel
 - accuracy: angle/area less accurate than line length
 - arclength also less accurate than line length



- polar area chart
 - area marks with length channel
 - more direct analog to bar charts



- data
 - 1 categ key attrib, 1 quant value attrib
- task
 - part-to-whole judgements

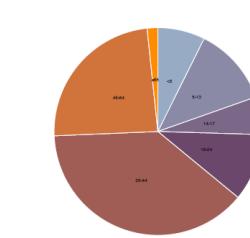
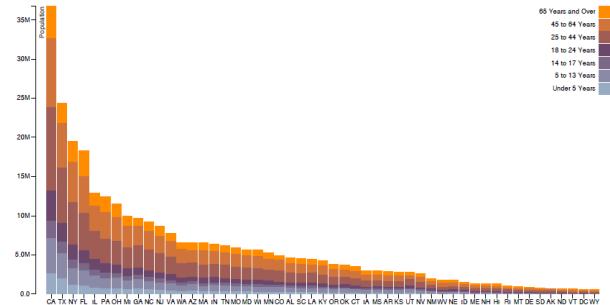
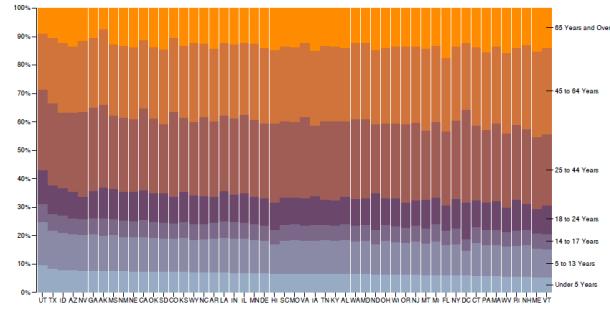


[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Idiom: Normalized Stacked Bar Chart

Idioms: **normalized stacked bar chart**

- task
 - part-to-whole judgements
- normalized stacked bar chart
 - stacked bar chart, normalized to full vert height
 - single stacked bar equivalent to full pie
 - high information density: requires narrow rectangle
- pie chart
 - information density: requires large circle



<http://bl.ocks.org/mbostock/3887235>,

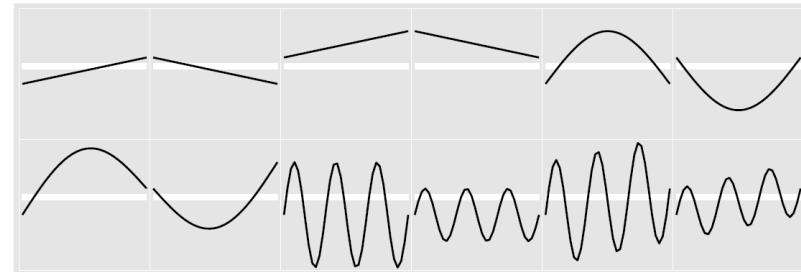
<http://bl.ocks.org/mbostock/3886208>,

<http://bl.ocks.org/mbostock/3886394>.

Idiom: GlyphMap

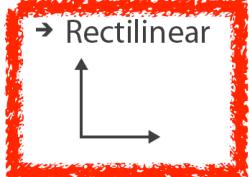
Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends

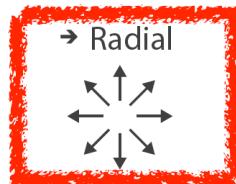


- radial good for cyclic patterns

④ Axis Orientation

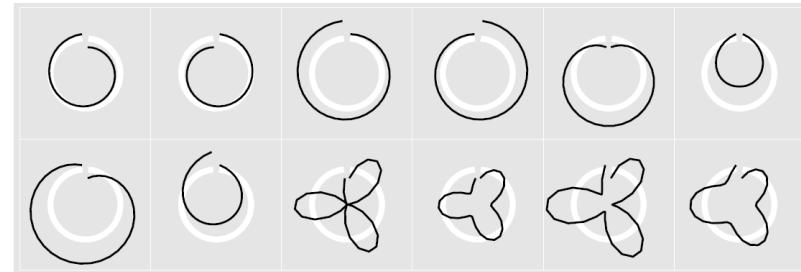


→ Rectilinear



→ Parallel

→ Radial



[*Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models*. Wickham, Hofmann, Wickham, and Cook. *Environmetrics* 23:5 (2012), 382–393.]

Tables Revisited

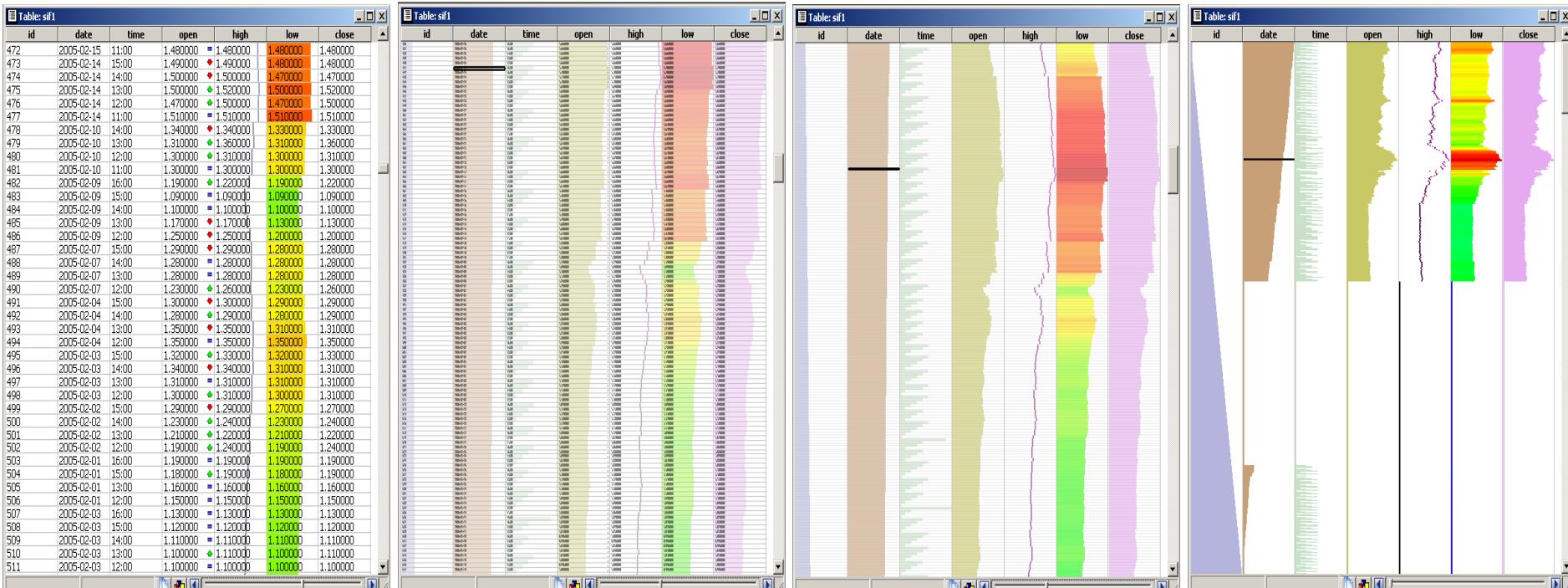
Enhancement: map values to colors / bar lengths

- numerical data: straightforward
- ordinal data: use order as value
- categorical data: define an order (arbitrary or application-dependent)

			date	time	stock	value			
id	category	name	date 1	date 2	time 3	open	high	low	close
1611	oil	SNP	2004-01-09	12:00		0.146200	0.146200	0.145200	0.145200
1610	oil	SNP	2004-01-09	13:00		0.145200	0.146200	0.145200	0.145200
1609	oil	SNP	2004-01-09	14:00		0.145200	0.145200	0.145200	0.145200
1608	oil	SNP	2004-01-09	15:00		0.145200	0.145200	0.144200	0.144200
1607	oil	SNP	2004-01-12	11:00		0.144200	0.144200	0.143200	0.143200
1606	oil	SNP	2004-01-12	12:00		0.143200	0.143200	0.142300	0.142300
1605	oil	SNP	2004-01-12	13:00		0.142300	0.142300	0.140300	0.141300
1604	oil	SNP	2004-01-12	14:00		0.140300	0.140300	0.140300	0.140300
1603	oil	SNP	2004-01-12	15:00		0.140300	0.141300	0.140300	0.141300
1602	oil	SNP	2004-01-13	11:00		0.141300	0.141300	0.140300	0.140300
1601	oil	SNP	2004-01-13	12:00		0.140300	0.142300	0.140300	0.141300
1600	oil	SNP	2004-01-13	13:00		0.141300	0.142300	0.141300	0.141300
1599	oil	SNP	2004-01-13	14:00		0.141300	0.142300	0.141300	0.142300
1598	oil	SNP	2004-01-13	15:00		0.141300	0.142300	0.141300	0.141300
1597	oil	SNP	2004-01-14	11:00		0.141300	0.141300	0.140300	0.140300
1596	oil	SNP	2004-01-14	12:00		0.141300	0.142300	0.141300	0.141300
1595	oil	SNP	2004-01-14	13:00		0.142300	0.143200	0.142300	0.142300
1594	oil	SNP	2004-01-14	14:00		0.142300	0.142300	0.141300	0.141300
1593	oil	SNP	2004-01-15	11:00		0.141300	0.142300	0.141300	0.141300
1592	oil	SNP	2004-01-15	12:00		0.141300	0.141300	0.141300	0.141300
1591	oil	SNP	2004-01-15	13:00		0.141300	0.141300	0.141300	0.141300
1590	oil	SNP	2004-01-15	14:00		0.141300	0.142300	0.141300	0.141300
1589	oil	SNP	2004-01-15	15:00		0.141300	0.142300	0.141300	0.142300
1588	oil	SNP	2004-01-16	11:00		0.141300	0.141300	0.140300	0.140300
1587	oil	SNP	2004-01-16	12:00		0.140300	0.141300	0.140300	0.141300
1586	oil	SNP	2004-01-16	13:00		0.140300	0.140300	0.140300	0.140300
1585	oil	SNP	2004-01-16	14:00		0.140300	0.141300	0.140300	0.140300
1584	oil	SNP	2004-01-16	15:00		0.140300	0.141300	0.140300	0.141300
896	oil	SNP	2004-07-29	14:00		0.860000	0.860000	0.860000	0.860000
895	oil	SNP	2004-07-29	15:00		0.855000	0.860000	0.855000	0.860000
894	oil	SNP	2004-07-30	12:00		0.860000	0.865000	0.860000	0.865000
893	oil	SNP	2004-07-30	13:00		0.860000	0.860000	0.860000	0.860000
892	oil	SNP	2004-07-30	14:00		0.860000	0.860000	0.860000	0.860000
891	oil	SNP	2004-07-30	15:00		0.860000	0.860000	0.860000	0.860000
890	oil	SNP	2004-08-02	12:00		0.865000	0.865000	0.865000	0.865000
889	oil	SNP	2004-08-02	13:00		0.870000	0.870000	0.865000	0.865000
888	oil	SNP	2004-08-02	14:00		0.870000	0.875000	0.870000	0.875000
887	oil	SNP	2004-08-02	15:00		0.875000	0.880000	0.875000	0.880000
886	oil	SNP	2004-08-03	11:00		0.875000	0.875000	0.875000	0.875000
885	oil	SNP	2004-08-03	12:00		0.875000	0.875000	0.875000	0.875000
884	oil	SNP	2004-08-03	13:00		0.875000	0.875000	0.875000	0.875000

Enhancement: the ‘table lens’ [Rao et al, ‘94]

- zoom the cells, but keep the layout
- fade out text, fade in colored / scaled bars
- → replace the entire table by a set of 1D graphs



text opacity=1
font size=12pt

text opacity ↓
font size ↓

text not drawn
bar opacity↑

bar opacity=1
simplification=on

min

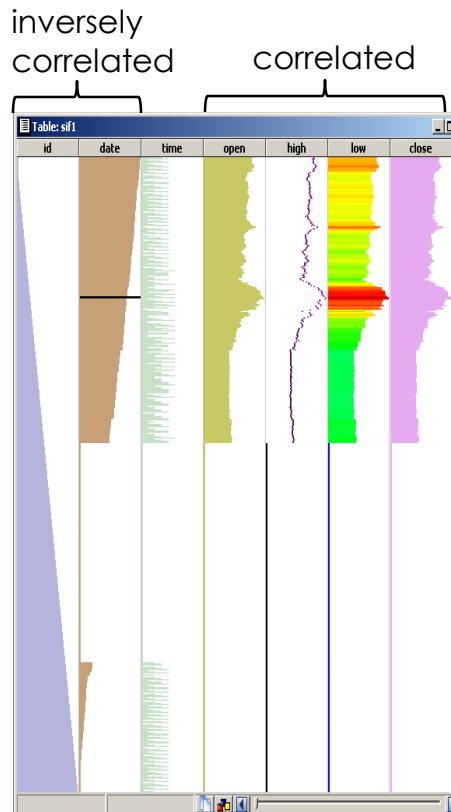
zoom level

max

Tables

Third enhancement: single-column sorting

- sort table on user-selected column value
- zoom out mode shows **distribution** and **correlation** of column values



Example: build cost analysis of a software system S

- rows = files f ; columns = cost metrics
- build impact (f) = cost to rebuild S if f is touched

