

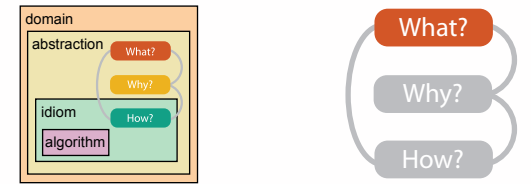
Week 2: Tasks & Data, Tables

Tamara Munzner
 Department of Computer Science
 University of British Columbia

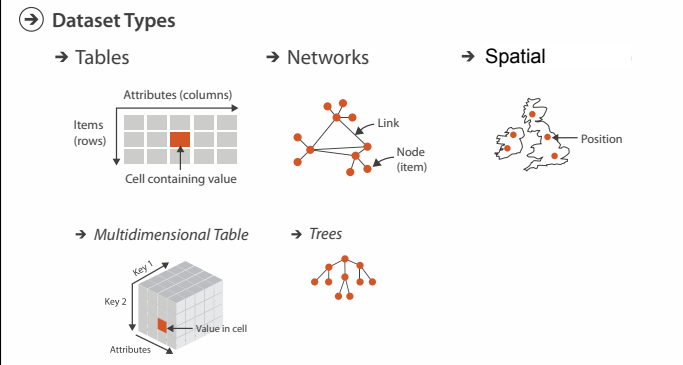
JRNL 520M, Special Topics in Contemporary Journalism: Visualization for Journalists
 Week 2: 22 September 2015

<http://www.cs.ubc.ca/~tmm/courses/journ15>

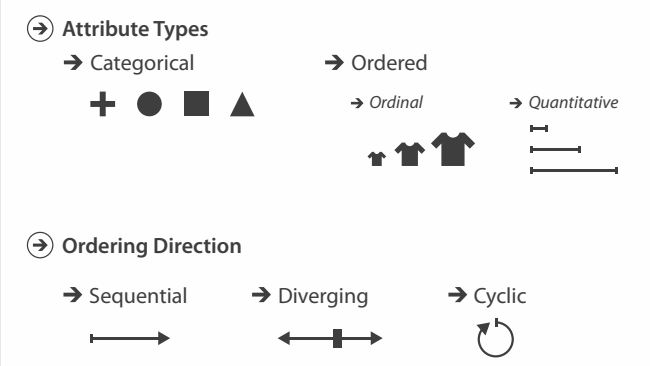
Data abstraction: What



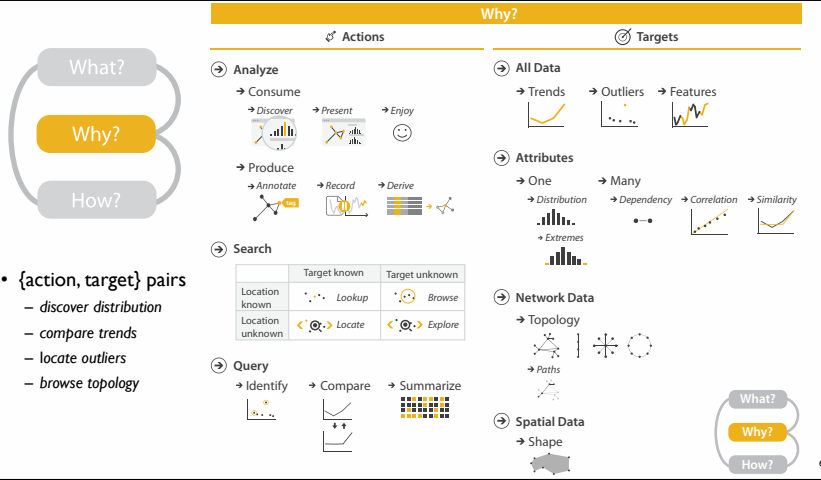
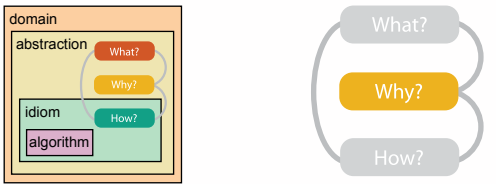
Data Abstraction



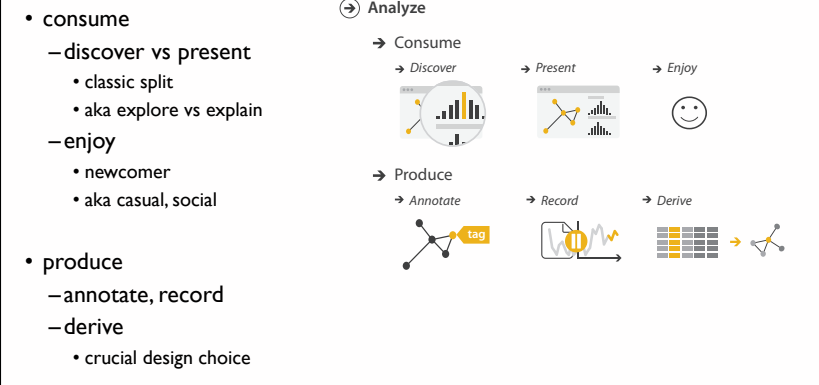
Attribute types



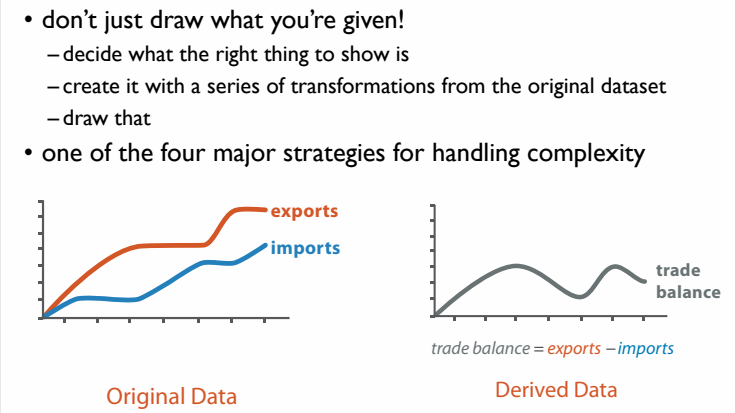
Tasks abstraction: Why



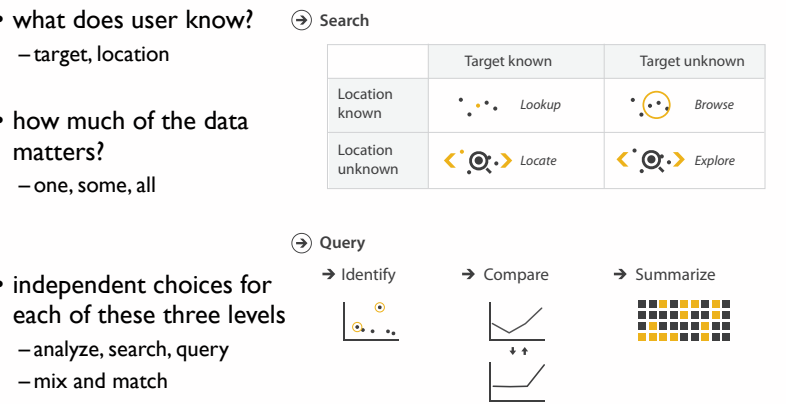
Actions: Analyze



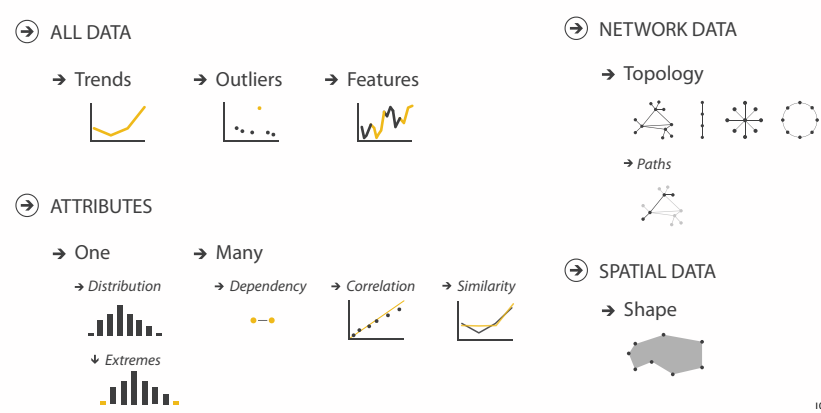
Derive



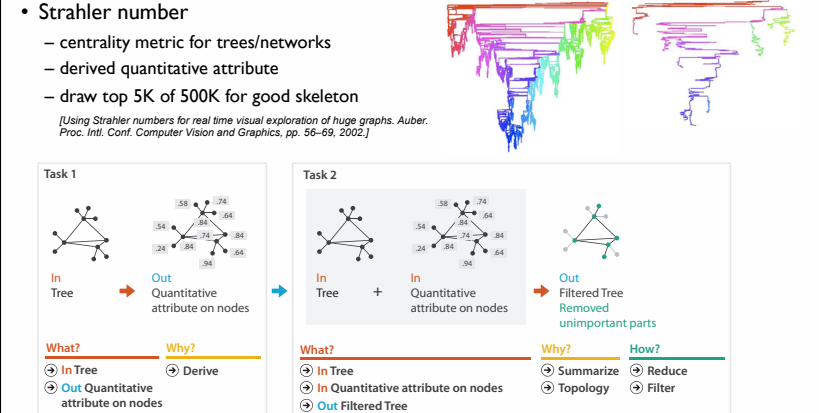
Actions: Search, query



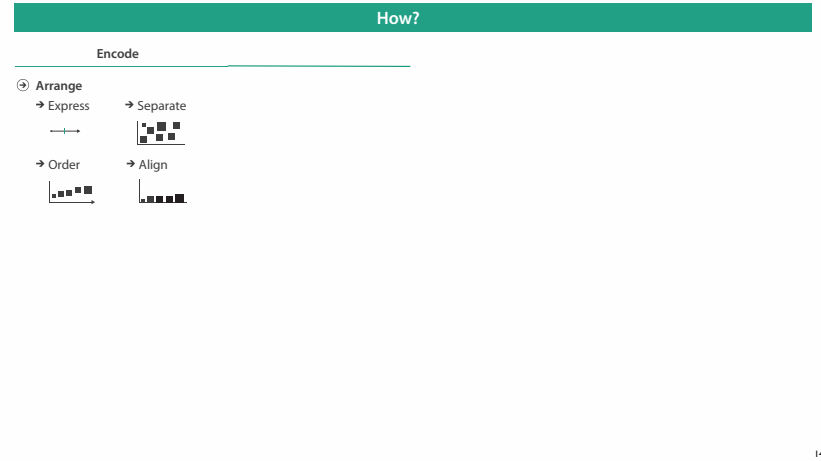
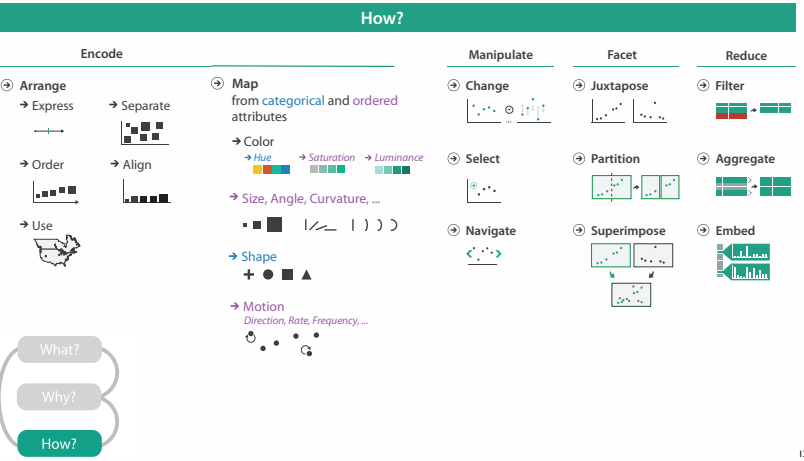
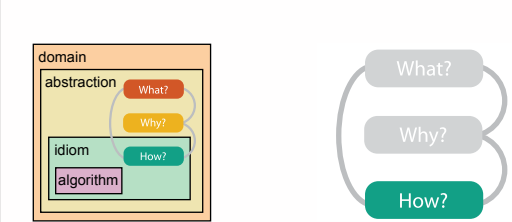
Why: Targets



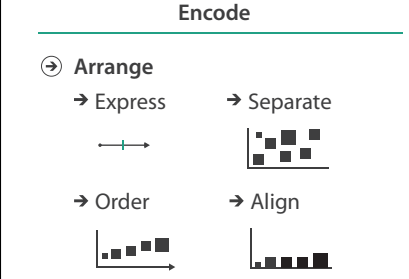
Analysis example: Derive one attribute



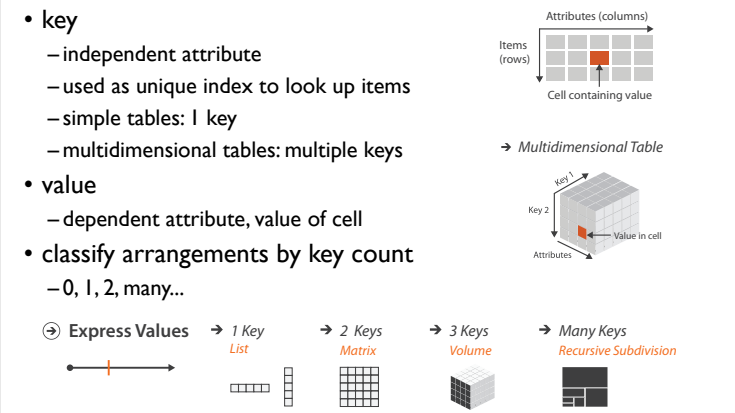
Visual encoding and interaction idiom: How



Encode tables: Arrange space



Keys and values



Idiom: scatterplot

Express Values

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

[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.] 17

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

18

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

19

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.] 20

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).] 21

Idiom: line chart

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

22

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 attribs
 - 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.] 23

Idiom: heatmap

- two keys, one value
 - data
 - 2 categ attribs (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
 - IM items, 100s of categ levels, ~10 quant attrib levels

→ 1 Key List → 2 Keys Matrix → Many Keys Recursive Subdivision 24

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

25

Axis Orientation

→ Rectilinear → Parallel → Radial

26

Idioms: scatterplot matrix, parallel coordinates

- scatterplot matrix (SPLOM)
 - rectilinear axes, point mark
 - all possible pairs of axes
 - scalability
 - one dozen attribs
 - 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 attribs
 - hundreds of items

	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/] 27

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

[Hyperdimensional Data Analysis: Using Parallel Coordinates. Wegman. Journ. American Statistical Association 85:411 (1990), 664–675.] Figure 3. Parallel Coordinate Plot of Six Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8, \text{ and } -1.$ 28

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

[Visman: Facilitating Risk Assessment and Decision Making in Fisheries Management. Booshehrian, Müller, Peterman, and Munzner. Technical Report TR 2011-04, Simon Fraser University, School of Computing Science, 2011.] 29

Idioms: pie chart, polar area chart

- pie chart
 - area marks with angle channel
 - accuracy: angle/area much 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.] 30

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/3886374 31

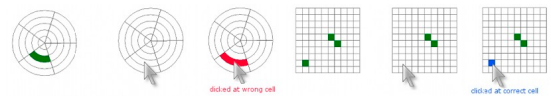
Idiom: glyphmaps

- rectilinear good for linear vs nonlinear trends
 - linear trends
 - nonlinear trends
- radial good for cyclic patterns
 - cyclic patterns

[Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. Wickham, Hofmann, Wickham, and Cook. Environmetrics 23:5 (2012), 382–393.] 32

Orientation limitations

- rectilinear: scalability wrt #axes
 - 2 axes best
 - 3 problematic
 - more in afternoon
 - 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
 - asymmetry: angles lower precision than lengths
 - sometimes can be exploited!



[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935-942, 2010.]

33

Axis Orientation

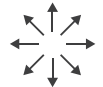
→ Rectilinear



→ Parallel



→ Radial

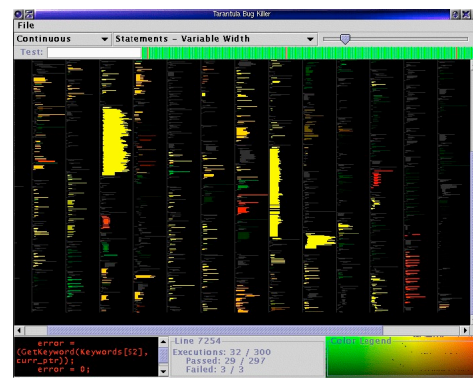


Layout Density

→ Dense



dense software overviews



[Visualization of test information to assist fault localization. Jones, Harrold, Staska. Proc. ICSE 2002, p. 467-477.]

34

Further reading

- Visualization Analysis and Design. Tamara Munzner. CRC Press, 2014.
 - Chap 2: Data Abstraction
 - Chap 3: Task Abstraction
 - Chap 7: Tables
- A Brief History of Data Visualization. Friendly. 2008.
<http://www.datavis.ca/milestones>

Now

- Break (15 min)
- Demo (30 min)
 - Guest lecture/demo from Robert Kosara on data wrangling
 - Tableau and Wrangler
- Lab 2 (45 min)

35

Lab/Assignment 2

- two main datasets
 - development aid from Guardian Datablog
 - your choice from small set
- focus on tasks and spatial layout as discussed in class for your exploration, story discovery, and writeup
 - provide rationale justifying your design decisions
- submit next week
 - by 9am Tue, email tmm@cs.ubc.ca with subject JOURN Week 2

37

36