Information Visualization

Tables

Tamara Munzner
Department of Computer Science
University of British Columbia

Week 3 async video, Jan 2021

https://www.students.cs.ubc.ca/~cs-436v/21Jan/
Tables
Focus on Tables

Dataset Types

- **Tables**
  - Attributes (columns)
  - Items (rows)
  - Cell containing value
  - Multidimensional Table

- **Networks**
  - Link
  - Node (item)
  - Trees

- **Spatial**
  - Fields (Continuous)
  - Geometry (Spatial)
  - Grid of positions
  - Cell
  - Attributes (columns)
  - Value in cell

- **Position**
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
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, ...

- **0 Keys**
  - Express Values

- **1 Key**
  - List

- **2 Keys**
  - Matrix
Idiom: scatterplot

- **express** values (magnitudes)
  - quantitative attributes
- no keys, only values
Idiom: **scatterplot**

- **express** values (magnitudes)
  - quantitative attributes
- **no keys, only values**
  - data
    - 2 quant attribs
  - mark: points
  - channels
    - horiz + vert position

[...]

Idiom: scatterplot

- **express** values (magnitudes)
  - 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

Scatterplots: Encoding more channels

• additional channels viable since using point marks
  – color
  – size (bubbleplots)
    • radius is misleading, take square root since area grows quadratically
  – shape

https://www.d3-graph-gallery.com/graph/bubble_basic.html
Scatterplot tasks
Scatterplot tasks

- correlation

https://www.mathsisfun.com/data/scatter-xy-plots.html
Scatterplot tasks

- correlation

- clusters/groups, and clusters vs classes

https://www.mathsisfun.com/data/scatter-xy-plots.html

https://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/
Some keys

- 0 Keys
- Express Values

- 1 Key
  - List

- 2 Keys
  - Matrix
Some keys: Categorical regions

- **Separate**

- **Order**

- **Align**

1 Key

2 Keys

3 Keys

Many Keys

List

Recursive Subdivision

Volume

Matrix

Rectilinear

Parallel

Radial
Some keys: Categorical regions

- **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
Separated but not aligned or ordered

• limitation: hard to make comparisons with size (vs aligned position)
Separated and aligned but not ordered

- limitation: hard to know rank. what's 4th? what's 7th?
Separated and aligned and ordered

- best case
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
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: asymmetric
    - for *stacked* key attrib, 10-12 levels [segments]
    - for *main* key attrib, dozens to hundreds of levels [bars]

https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html
Idiom: **streamgraph**

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
- data
  - 1 categorical key attrib (movies)
  - 1 ordered key attrib (time)
  - 1 quantitative value attrib (counts)
- derived data
  - geometry: layers, where height encodes counts
  - 1 quantitative attrib (layer ordering)

Idiom: **streamgraph**

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 category key attrib (movies)
    - 1 ordered key attrib (time)
    - 1 quantitative attrib (counts)
- derived data
  - geometry: layers, where height encodes counts
  - 1 quantitative attrib (layer ordering)
- scalability
  - hundreds of time keys
  - dozens to hundreds of movies keys
    - more than stacked bars: most layers don’t extend across whole chart
Idiom: **dot / line chart**

• one key, one value
  – data
    • 2 quant attrs
  – mark: points
    AND line connection marks between them
  – channels
    • aligned lengths to express quant value
    • separated and ordered by key attrib into horizontal regions
Idiom: **dot / line chart**

- one key, one value
  - data
    - 2 quant attribs
  - mark: points
    AND 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
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”

Chart axes: label them!

• best practice to label
  – few exceptions: individual small multiple views could share axis label

https://xkcd.com/833/
Chart axes: avoid cropping y axis

- include 0 at bottom left or slope misleads

[Truncating the Y-Axis: Threat or Menace?
Correll, Bertini, & Franconeri, CHI 2020.]
Chart axes: avoid cropping y axis

- include 0 at bottom left or slope misleads
  - some exceptions (arbitrary 0, small change matters)

[Truncating the Y-Axis: Threat or Menace? Correll, Bertini, & Franconeri, CHI 2020.]
Idiom: **dual-axis line charts**

- controversial
  - acceptable if commensurate
  - beware, very easy to mislead!

Dual axis charts: case study

misleading original

first redesign: commensurate axes

second redesign: additional data for context

http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html
Idiom: Indexed line charts

• data: 2 quant attrs
  – 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/CAStateRevenues/Revenues
Idiom: **Gantt charts**

- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
  - mark: line
    - length: duration
- channels
  - horiz position: start time (+end from duration)
- task
  - emphasize temporal overlaps & start/end dependencies between items
- scalability
  - dozens of key levels [bars]
  - hundreds of value levels [durations]
Idiom: Slopegraphs

- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
  - mark: point + line
    - line connecting mark between pts
- channels
  - 2 vertical pos: express attrib value
  - (linewidth/size, color)
- task
  - emphasize changes in rank/value
- scalability
  - hundreds of value levels
  - dozens of items

https://public.tableau.com/profile/ben.jones#!/vizhome/Slopegraphs/Slopegraphs
2 Keys

- Express Values
- Arrangement of Tables
- Expression of Values
- Separation, Ordering, Alignment of Regions
- Axis Orientation
- Layout Density
  - Dense
  - Space-filling
- Rectilinear
  - Parallel
  - Radial
- Recursive Subdivision
- Volume
- Matrix
Idiom: heatmap

- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: point
    - 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
Heatmap reordering

https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html
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
  - task: assess quality of clusters found by automatic methods
Axis Orientation

- Rectilinear
- Parallel
- Radial
Idioms: radial bar chart, star plot

• star plot
  – line mark, radial axes meet at central point
• radial bar chart
  – line mark, radial axes meet at central ring
  – channels: length, angle/orientation
• bar chart
  – rectilinear axes, aligned vertically

• accuracy
  – length not aligned with radial layouts
    • less accurately perceived than rectilinear aligned

Idiom: radar plot

• radial line chart
  – point marks, radial layout
  – connecting line marks

• avoid unless data is cyclic
"Radar graphs: Avoid them (99.9% of the time)"

Idioms: pie chart, coxcomb chart

• pie chart
  – line marks with angle channel: variable (sector) width
    • separated & ordered radially, uniform height
  – accuracy: less accurate than rectilinear aligned line length

• coxcomb chart
  – line marks with length channel: variable length
    • separated & ordered radially, uniform width
  – more direct analog to bar charts

• data
  – 1 categ key attrib, 1 quant value attrib

• task
  – part-to-whole judgements

Coxcomb / nightingale rose / polar area chart

- invented by Florence Nightingale:
  Diagram of the Causes of Mortality in the Army in the East
Coxcomb: perception

- encode: angle, length
- decode/perceive: area

- nonuniform line/sector width, so area variation is nonlinear wrt line mark length!

- rectilinear bar chart safer: uniform width so area is linear with line mark length
Pie charts: perception

• some empirical evidence that people respond to arc length
  – decode/perceive: not angles
  – maybe also areas?…
• donut charts no worse than pie charts


Pie charts: best practices

• not so bad for two (or few) levels, for part-to-whole task

https://eagereyes.org/pie-charts
Pie charts: best practices

- not so bad for two (or few) levels, for part-to-whole task
- dubious for several levels if details matter

https://eagereyes.org/pie-charts
Pie charts: best practices

• not so bad for two (or few) levels, for part-to-whole task
• dubious for several levels if details matter
• terrible for many levels

https://eagereyes.org/pie-charts
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/3886208
http://bl.ocks.org/mbostock/3887235
http://bl.ocks.org/mbostock/3886394
Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends

- radial good for cyclic patterns
  - evaluating periodicity

Axis Orientation

- Rectilinear
- Parallel
- Radial
Idiom: **SPLOM**

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

Wilkinson et al., 2005
Idioms: parallel coordinates

- scatterplot limitation
  - visual representation with orthogonal axes
  - can show only two attributes with spatial position channel
**Idioms: 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**
  - parallel axes, jagged line for item
  - rectilinear axes, item as point
    - axis ordering is major challenge
  - scalability
    - dozens of attribs
    - hundreds of items

---

**Table**

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Physics</th>
<th>Dance</th>
<th>Drama</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>85</td>
<td>95</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>Physics</td>
<td>90</td>
<td>80</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Dance</td>
<td>65</td>
<td>50</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Drama</td>
<td>50</td>
<td>40</td>
<td>95</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
<td>80</td>
<td>90</td>
</tr>
</tbody>
</table>

*after [Visualization Course Figures. McGuffin, 2014.](http://www.michaelmcguffin.com/courses/vis/)*
Task: Correlation

- scatterplot matrix
  - positive correlation
    - diagonal low-to-high
  - negative correlation
    - diagonal high-to-low
  - uncorrelated: spread out

- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings


https://www.mathsisfun.com/data/scatter-x-y-plots.html
Parallel coordinates, limitations

• 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 of interaction: human-powered search
  – some algorithms proposed, none fully solve
Orientation limitations

• rectilinear: scalability wrt #axes
  • 2 axes best, 3 problematic, 4+ impossible
Orientation limitations

• rectilinear: scalability wrt #axes
  • 2 axes best, 3 problematic, 4+ impossible
• parallel: unfamiliarity, training time
Orientation limitations

- rectilinear: scalability wrt #axes
  - 2 axes best, 3 problematic, 4+ impossible
- parallel: unfamiliarity, training time
- radial: perceptual limits
  - polar coordinate asymmetry
    - angles lower precision than length
    - nonuniform sector width/size depending on radial distance
  - frequently problematic
    - but sometimes can be deliberately exploited!
      - for 2 attrs of very unequal importance

Layout Density

- Dense
  - Illustration of dense layout density

- Space-Filling
  - Illustration of space-filling layout density
Idiom: Dense software overviews

- **data**: text
  - text + 1 quant attrib per line

- **derived data**:
  - one pixel high line
  - length according to original

- **color line by attrib**

- **scalability**
  - 10K+ lines

Encode tables: Arrange space

Encode

↺ Arrange
↺ Express
↺ Separate
↺ Order
↺ Align
Arrange tables

Express Values

Separate, Order, Align Regions

- Separate
- Order
- Align

1 Key
List

2 Keys
Matrix

Axis Orientation

- Rectilinear
- Parallel
- Radial

Layout Density

- Dense

Rectilinear
Parallel
Radial
### How?

<table>
<thead>
<tr>
<th>Encode</th>
<th>Manipulate</th>
<th>Facet</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arrange</strong></td>
<td><strong>Map</strong></td>
<td><strong>Juxtapose</strong></td>
<td><strong>Filter</strong></td>
</tr>
<tr>
<td>Express</td>
<td>from <strong>categorical</strong> and <strong>ordered</strong> attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separate</td>
<td><strong>Color</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td>Hue</td>
<td><strong>Select</strong></td>
<td><strong>Aggregate</strong></td>
</tr>
<tr>
<td>Align</td>
<td>Saturation</td>
<td>Partition</td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td>Luminance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Size, Angle, Curvature, ...</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Shape</strong></td>
<td><strong>Navigate</strong></td>
<td>Embed</td>
</tr>
<tr>
<td></td>
<td>+ ○ ■ △</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Motion</strong></td>
<td>Superimpose</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Direction, Rate, Frequency, ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>○ ○ ○</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Credits

• Visualization Analysis and Design (Ch 7)
• Alex Lex & Miriah Meyer, http://dataviscourse.net/
• Ben Jones, UW/Tableau