Information Visualization
Tables

Tamara Munzner
Department of Computer Science
University of British Columbia

Lect 6/7, 23/28 Jan 2020

https://www.cs.ubc.ca/~tmm/courses/436V-20
Tables
Focus on Tables

Dataset Types

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

- **Multidimensional Table**
  - Attributes
  - Key 1
  - Key 2
  - Value in cell

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

- **Trees**
  - Value in cell

- **Spatial**

- **Fields (Continuous)**

- **Geometry (Spatial)**
  - Position
  - Grid of positions
  - Attributes (columns)
  - Value in cell
Exercise: Sketch 2 ways to visualize each table

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Best 100 m</th>
<th>Furthest Jump</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amy</td>
<td>16</td>
<td>13.2</td>
<td>5.2</td>
<td>F</td>
</tr>
<tr>
<td>Basil</td>
<td>18</td>
<td>12.4</td>
<td>4.2</td>
<td>F</td>
</tr>
<tr>
<td>Clara</td>
<td>14</td>
<td>14.1</td>
<td>2.5</td>
<td>F</td>
</tr>
<tr>
<td>Desmond</td>
<td>22</td>
<td>10.01</td>
<td>6.3</td>
<td>M</td>
</tr>
<tr>
<td>Charles</td>
<td>19</td>
<td>11.3</td>
<td>5.3</td>
<td>M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BPM T1</th>
<th>BPM T2</th>
<th>BPM T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amy</td>
<td>90</td>
<td>130</td>
<td>150</td>
</tr>
<tr>
<td>Basil</td>
<td>70</td>
<td>110</td>
<td>109</td>
</tr>
<tr>
<td>Clara</td>
<td>60</td>
<td>140</td>
<td>141</td>
</tr>
<tr>
<td>Desmond</td>
<td>84</td>
<td>100</td>
<td>108</td>
</tr>
<tr>
<td>Charles</td>
<td>81</td>
<td>110</td>
<td>130</td>
</tr>
</tbody>
</table>

• socrative: answer when done
Tackling tables

• homogeneity
  – same data type? same scales?

• need different approaches based on scale
  – how many attributes?
    • up to ~50: tractable with direct visual encoding
    • thousands: need transformations / analytical methods
  – how many items?
    • up to 1K: tractable with direct visual encoding
    • >> 10K: need transformations / analytical methods

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Gender</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>25</td>
<td>M</td>
<td>181</td>
</tr>
<tr>
<td>Alice</td>
<td>22</td>
<td>F</td>
<td>185</td>
</tr>
<tr>
<td>Chris</td>
<td>19</td>
<td>M</td>
<td>175</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BPM 1</th>
<th>BPM 2</th>
<th>BPM 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>65</td>
<td>120</td>
<td>145</td>
</tr>
<tr>
<td>Alice</td>
<td>80</td>
<td>135</td>
<td>185</td>
</tr>
<tr>
<td>Chris</td>
<td>45</td>
<td>115</td>
<td>135</td>
</tr>
</tbody>
</table>
Analytic component

Scatterplot Matrices [Bostock]

Parallel Coordinates [Bostock]

Pixel-based visualizations / heat maps

Multidimensional Scaling [Doerk 2011]

no / little analytics

strong analytics component
Tasks and techniques

Magnitude  Distribution  Deviation  Correlation
Ranking    Part to whole  Change over Time

https://github.com/ft-interactive/chart-doctor/tree/master/visual-vocabulary
https://gramener.github.io/visual-vocabulary-vega/#/Magnitude/
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...

Express Values

- **1 Key**
  - List
- **2 Keys**
  - Matrix
- **3 Keys**
  - Volume
- **Many Keys**
  - Recursive Subdivision
0 Keys: Express values (magnitudes)
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

Scatterplots: Encoding more channels

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

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

Express Values

1 Key
List

2 Keys
Matrix

3 Keys
Volume

Many Keys
Recursive Subdivision
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
- **3 Keys**: Volume
- **Many Keys**: Recursive Subdivision
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
LIMITATION: Hard to know rank. What’s the 4th most? The 7th?
Separated but not Ordered or Aligned

LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]
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

[Link to stacked bar chart example](https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html)
Idiom: streamgraph

- generalized stacked graph
  - emphasizing horizontal continuity
    - vs vertical items
  - data
    - 1 categ key attrib (movies)
    - 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 movies keys
    - more than stacked bars, since most layers don’t extend across whole chart


**Idiom: dot plot / 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

- labelled axis is critical
- avoid cropping y-axis
  - include 0 at bottom left
  - or slope misleads

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

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

**Idiom: connected scatterplots**

- scatterplot with line connection marks
  - popular in journalism
  - horiz + vert axes: value attribs
  - line connection marks: temporal order
  - alternative to dual-axis charts
    - horiz: time
    - vert: two value attribs

- empirical study
  - engaging, but correlation unclear


http://steveharoz.com/research/connected_scatterplot/
Choosing line chart aspect ratios

• 1: banking to 45 (1980s)
  – Cleveland perceptual argument: most accurate angle judgement at 45

Fig 7.1 Sunspot Data: Aspect Ratio 1

Fig 7.2 Annual Report: Aspect Ratio 2
Choosing line chart aspect ratios

- 2: multi scale banking to 45 (2006)
  - frequency domain analysis to find ratios
    - FFT the data, convolve with Gaussian to smooth
    - find interesting spikes/ranges in power spectrum
  - cull nearby regions if similar, ensure overview
  - create trend curves (red) for each aspect ratio

Sunspot Cycles
Aspect Ratio = 3.96

[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006]
Choosing line chart aspect ratios

• 3: arc length based aspect ratio (2011)
  – minimize the arc length of curve while keeping the area of the plot constant
  – parametrization and scale invariant
  – symmetry preserving
  – robust & fast to compute

• meta-points from this progression
  – young field; prescriptive advice changes rapidly
  – reasonable defaults required deep dive into perception meets math

[Arc Length-Based Aspect Ratio Selection. Talbot, Gerth, and Hanrahan. Proc InfoVis 2011]
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/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
  - hundreds of value levels


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

https://public.tableau.com/profile/ben.jones#!/vizhome/Slopegraphs/Slopegraphs
Breaking conventions

• presentation vs exploration
  – engaging/evocative
  – inverted y axis
  • blood drips down on Poe

[Slide inspired by Ben Jones]
2 Keys

Express Values → 1 Key
List

Express Values → 2 Keys
Matrix

Express Values → 3 Keys
Volume

Express Values → Many Keys
Recursive Subdivision

ARRANGE TABLES
EXPRESS VALUES
SEPARATE, ORDER, ALIGN REGIONS
AXIS ORIENTATION
LAYOUT DENSITY
Dense Spacefilling
Separate Order Align
1 Key 2  Keys 3 Keys Many Keys
List Recursive SubdivisionVolumeMatrix
Rectilinear Parallel Radial
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
    - 1M items, 100s of categ levels, ~10 quant attrib levels
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

• 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

Radial Orientation: Radar Plots

LIMITATION: Not good when categories aren’t cyclic

[Slide courtesy of Ben Jones]
"Diagram of the causes of mortality in the army in the East" (1858)

[Slide courtesy of Ben Jones]
“Radar graphs: Avoid them (99.9% of the time)”

[Slide courtesy of Ben Jones]

Idioms: **pie chart, polar area chart**

- **pie chart**
  - line marks with angle channel: variable (sector) width
  - separated & aligned radially, uniform height
  - perceived: probably not angle! maybe area or arc length
  - accuracy: all are less accurate than line length

- **polar area chart**
  - line marks with length channel: variable length
  - separated & aligned radially, uniform width
  - more direct analog to bar charts

- **data**
  - 1 categ key attrib, 1 quant value attrib

- **task**
  - part-to-whole judgements

Pie chart perception

• some empirical evidence that people respond to arc length
  – not angles
  – maybe also areas?…

• donut charts no worse than pie charts


Pie chart best practices

• not 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

---


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
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


Orientation limitations

- **rectilinear**: scalability wrt #axes
  - 2 axes best
  - 3 problematic
  - 4+ impossible
- **parallel**: unfamiliarity, training time

**Axis Orientation**
- **Rectilinear**
- **Parallel**
- **Radial**
Radial orientation

• perceptual limits
  – polar coordinate asymmetry
    • angles lower precision than lengths
    • frequently problematic
    • sometimes can be deliberately exploited!
      • for 2 attribs of very unequal importance

Layout density

Layout Density

- Dense
- Space-Filling
Idiom: Dense software overviews

- Layout Density
  - Dense
    - 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

Why?

How?

What?
Arrange tables

Express Values

Separate, Order, Align Regions

Separate

Order

Align

Axis Orientation

Rectilinear

Parallel

Radial

Layout Density

Dense

Space-Filling

1 Key

List

2 Keys

Matrix

3 Keys

Volume

Many Keys

Recursive Subdivision
### 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>Change</strong></td>
<td><strong>Juxtapose</strong></td>
<td><strong>Filter</strong></td>
</tr>
<tr>
<td>➔ Express</td>
<td>➔ Hue</td>
<td>➔ Partition</td>
<td>➔ Aggregate</td>
</tr>
<tr>
<td>➔ Order</td>
<td>➔ Saturation</td>
<td>➔ Superimpose</td>
<td>➔ Embed</td>
</tr>
<tr>
<td>➔ Use</td>
<td>➔ Luminance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Map** from categorical and ordered attributes
  - **Shape**
    - + ○ ● ■ ▲
  - **Size, Angle, Curvature, ...**
  - **Motion**
    - Direction, Rate, Frequency, ...

- **Use**
- **Filter**
- ** Aggregate**
- **Embed**
Upcoming

• D3 videos to watch this week
  – Making a Bar Chart with D3 and SVG [30 min]
• Quiz 3, due by Fri Jan 24, 8am
• Programming Exercise 1, due Wed Jan 29
Credits

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