Week 2: Arrange Tables

Tamara Munzner
Department of Computer Science
University of British Columbia

JRNL 520H, Special Topics in Contemporary Journalism: Data Visualization
Week 2: 20 September 2016

http://www.cs.ubc.ca/~tmm/courses/journ16
Finding us

• office hours in Sing Tao bldg
  – 1-ish to 3-ish pm Tuesdays in Room 313: Tamara and/or Caitlin
  – by appointment: Tamara in ICICS/CS bldg Room X661

• email other times
  – tmm@cs.ubc.ca, caitlin@discoursemedia.org

• course page is font of all information
  – don’t forget to refresh, frequent updates
  – http://www.cs.ubc.ca/~tmm/courses/journ16
Last Time
Demo 1: Basic Visual Encoding & Dashboarding

• Tableau Lessons
  – Dimensions (categorical) and Measures (quantitative)
  – drag and drop to create visual encodings
  – combining multiple charts side by side into dashboards

• Big Ideas
  – see different patterns with different visual encodings
Demo 2: Vancouver Election Results

• Tableau Lessons
  – sorting along axis
  – disaggregate into multiple charts

• Big Ideas
  – absolute numbers can sometimes mislead
  – check hunches with relative percentages!
Demo 3: Vancouver Crime

• Tableau Lessons
  – multiple pills on a shelf, pill ordering
  – show filters
  – undo
  – duplicate & rename tabs

• Big Ideas
  – underlying causes can be tricky to understand
Arrange Tables
### Encode

<table>
<thead>
<tr>
<th>Arrange</th>
<th>Express</th>
<th>Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>Align</td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Map**
  - from *categorical* and *ordered* attributes
- **Color**
  - *Hue*, *Saturation*, *Luminance*
- **Size, Angle, Curvature, ...**
- **Shape**
  - + • ■ ▲
- **Motion**
  - *Direction, Rate, Frequency, ...*

### Manipulate

<table>
<thead>
<tr>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Change Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image2" alt="Select Diagram" /></td>
</tr>
</tbody>
</table>

### Facet

<table>
<thead>
<tr>
<th>Juxtapose</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Juxtapose Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partition</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4" alt="Partition Diagram" /></td>
</tr>
</tbody>
</table>

### Reduce

<table>
<thead>
<tr>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Filter Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aggregate</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image6" alt="Aggregate Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Superimpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Superimpose Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embed</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image8" alt="Embed Diagram" /></td>
</tr>
</tbody>
</table>
How?

Encode

Arrange

→ Express

→ Order

→ Separate

→ Align

How?

Encode Manipulate Facet Reduce
Encode tables: Arrange space

Encode

Arrange
- Express
- Separate

Order
- Align
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

Tables

Multidimensional Table
Idiom: **scatterplot**

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

[Express Values](#)

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

<table>
<thead>
<tr>
<th>1 Key</th>
<th>2 Keys</th>
<th>3 Keys</th>
<th>Many Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>List</td>
<td>Matrix</td>
<td>Volume</td>
<td>Recursive Subdivision</td>
</tr>
</tbody>
</table>
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 but not Ordered orAligned

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

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

**Idiom: line chart**

- 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 attr 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
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”

**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 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
Axis Orientation

→ Rectilinear

→ Parallel

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


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

Idioms: \textbf{normalized stacked bar chart}

- \textit{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

- \textit{pie chart}
  - information density: requires large circle

\texttt{http://bl.ocks.org/mbostock/3887235},
\texttt{http://bl.ocks.org/mbostock/3886208},
\texttt{http://bl.ocks.org/mbostock/3886394}.
Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends

- radial good for cyclic patterns

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

Layout Density

Dense

Basic Timelines – Working with Dates

- Yearly continuous
- Monthly
- Weekly
- Daily

[Slide courtesy of Ben Jones]
Column Charts

F. Scott Fitzgerald’s Earnings

[Slide courtesy of Ben Jones]
Inverted Column Charts

The Remarkable Career of Edgar Allan Poe

150 works

Source: https://en.wikipedia.org/wiki/Edgar_Allan_Poe_bibliography

[Slide courtesy of Ben Jones]
Slopegraphs

Barclay’s Premier League Tables: Comparing 2012/2013 Starts to 2013/2014 Starts

Select:
- Points

Clubs to include:
- (All)

2012/2013
- Manchester United, 36
- Manchester City, 33
- West Bromwich Albion, 26
- Everton, 22
- West Ham United, 22
- Arsenal, 21
- Liverpool, 19
- Fulham, 17
- Newcastle United, 14
- Sunderland, 13
- Southampton, 12

2013/2014
- Arsenal, 35
- Liverpool, 30
- Manchester City, 29
- Everton, 28
- Tottenham Hotspur, 27
- Newcastle United, 26
- Southampton, 22
- Manchester United, 22
- Aston Villa, 19
- Stoke City, 17
- West Bromwich Albion, 15
- West Ham United, 13
- Sunderland, 8

Data: 
Ref: 
By: 

[Slide courtesy of Ben Jones]
Connected Scatterplots

MLB Stats Over Time: Scatterplots vs. Dual Axes

Choose Variable 1
Number of Pitchers

Choose Variable 2
Strikeouts

Select a Year Range
1981
2012

Method #1. The Connected Scatterplot

[Slide courtesy of Ben Jones]
Dual Axis Line Plots


[Slide courtesy of Ben Jones]
Next

• Break (15 min)

• Demos (45 min)
  – Caitlin will walk through Tableau demos
  – you follow along step by step on your own laptop
  – Tamara will rove the room to help out folks who get stuck

• Lab (30 min)
  – you’ll get started on Tableau assignment