Information Visualization

Tables, LineUp, Bertifier

In Class: Pitches

Tamara Munzner
Department of Computer Science
University of British Columbia

Plan for today
• Pitches
  – 70-80 min pitches (2 min each)
  – up to 2 min Q&A after each
  – (break in the middle)
• remaining time: Reading Q&A
  – Tables
  – LineUp
  – Bertifier

Next week
• to read & discuss (async, before next class)
  – Vad book, Ch 10: Color
  – paper: Array Vie [design study & quant evaluation]
  – paper: Randows Revised [evaluation]

Q&A / Backup Slides

Focus on Tables

Dataset Types

• Items
  – dependent attribute, value of cell
  – mark: points
  – used as unique index to look up items
  – 2 quant attribs
  – grid of positions
  – horiz + vert 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
  – classify arrangements by keys used
  – 0, 1, 2...

Keys and values

• key
• dependent attribute
• no keys, only values

Scatterplot tasks

• correlation

Scatterplot tasks

• correlation
• clusters/groups, and clusters vs classes

Some keys

Scatterplots: Encoding more channels

• additional channels viable since using point marks
• color
• size (1 quart attribute, used to control 2D area)
• note radius would mislead, take square root since area grows quadratically
• shape

Scatterplot tasks

• express values (magnitudes)
  – quantitative attributes
  – no keys, only values
  – data
  – 2 quant attribs
  – mark points
  – channels
  – horiz = var position
  – tasks
  – find trends, outliers, distribution, correlation, clusters
  – scalability
  – hundreds of items

Keys and values

• key
• independent attribute
• no keys, only values

Idiom: scatterplot

• express values (magnitudes)
  – quantitative attributes
  – no keys, only values

Idiom: scatterplot

• express values (magnitudes)
  – quantitative attributes
  – no keys, only values

Idiom: scatterplot

• express values (magnitudes)
  – quantitative attributes
  – no keys, only values

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

Some keys

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

http://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/

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

https://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/
**Idiom: Gantt charts**
- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
    - mark: line
    - length: duration
    - channels
      - horizontal axis: start time
    - task
      - emphasis temporal overlaps & start/end dependencies between tasks
        - scalability
      - dozens of key levels (bars)
      - dozens of value levels
      - numerous items
  - rectilinear axes, aligned vertically
  - task mark: line length!

**Idiom: radar plot**
- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
    - mark: point + line
    - connecting mark between pts
    - channels
      - 3 verticall pos express attrib value
        - (width/height/color)
    - task
      - emphasis changes in rank/value
        - scalability
      - dozens of value levels
      - dozens of items

**Idiom: radar plot**
- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
    - mark: point + line
    - connecting mark between pts
    - channels
      - 3 verticall pos express attrib value
        - (width/height/color)
    - task
      - emphasis changes in rank/value
        - scalability
      - dozens of value levels
      - dozens of items

**Idiom: heatmap**
- two keys, one value
  - data
    - 2 categ attribs (gen, experimental condition)
  - 1 quant attrib (expression level)
  - marks: point
    - position & size in 3D space
      - indexed by 2 categ attribs
  - channels
    - color by quant attrib
      - (ordered diverging colormap)
    - task
      - find clusters, outliers
        - scalability
      - dozens of key levels (bars)
      - hundreds of value levels
      - dozens of items

**Idiom: Slopegraphs**
- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
    - mark: point + line
    - connecting mark between pts
    - channels
      - 3 verticall pos express attrib value
        - (width/height/color)
    - task
      - emphasis changes in rank/value
        - scalability
      - dozens of value levels
      - dozens of items

**Idiom: cluster heatmap**
- in addition
  - derived data
    - 2cluster 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

**Idiom: Gantt charts**
- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
    - mark: line
    - length: duration
    - channels
      - horizontal axis: start time
    - task
      - emphasis temporal overlaps & start/end dependencies between tasks
        - scalability
      - dozens of key levels (bars)
      - dozens of value levels
      - numerous items
  - rectilinear axes, aligned vertically
  - task mark: line length!

**Idiom: radar plot**
- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
    - mark: point + line
    - connecting mark between pts
    - channels
      - 3 verticall pos express attrib value
        - (width/height/color)
    - task
      - emphasis changes in rank/value
        - scalability
      - dozens of value levels
      - dozens of items

**Idiom: cluster heatmap**
- in addition
  - derived data
    - 2cluster 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

**Idiom: Gantt charts**
- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
    - mark: line
    - length: duration
    - channels
      - horizontal axis: start time
    - task
      - emphasis temporal overlaps & start/end dependencies between tasks
        - scalability
      - dozens of key levels (bars)
      - dozens of value levels
      - numerous items
  - rectilinear axes, aligned vertically
  - task mark: line length!

[Diagram of the Causes of Mortality in the Army in the East]
Pie charts: best practices
- not so bad for two (or fewer) levels, for part-to-whole task
- dubious for several levels if details matter
- terrible for many levels

Axis Orientation
- Rectilinear
- Parallel
- Radial

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

Idioms: glyph maps
- rectilinear good for linear vs nonlinear trends
- radial good for cyclic patterns
  - evaluating periodicity

Idioms: parallel coordinates
- scatterplot limitation
  - visual representation with orthogonal axes
  - can show only two attributes with spatial position channel
- alternate line up axes in parallel to show many attributes with position
  - zoom enabled with line by line segments
  - n is the number of attributes shown
- parallel coordinates
  - parallel axes, jagged line for ties
  - recalculate axes, zoom to point
  - axis ordering a major challenge
  - scalability
  - classes of artifact
  - hundreds of artifacts

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 crossing

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

Parallel coordinates, limitations
- visible patterns only between neighboring axis pairs
- how to pick axis order?
  - usual solution: reasonable axes, interactive exploration
  - some weakness as many other techniques
  - downside of interaction human-powered search
  - some algorithms proposed, none fully solve

Layout density
- Layout Density
- Space-Filling

Idioms: dense software overviews
- data/text
  - tags + 1 quadrant per line
- derived data:
  - one pixel high line
  - length according to original
- color line by artifact
- scalability
  - 10K+ lines

Arrange tables
- Express Values
  - Separate, Order, Align Regions
  - Separate
  - Order
  - Align
  - Layout Density

https://bl.ocks.org/mbostock/3886394