Tackling tables

- homogeneity
  - same data type? same scales?
- need different approaches based on scale
  - how many attributes?
  - up to ~50 translatable with direct visual encoding
  - thousands: need transformations / analytical methods
  - how many items?
  - up to ~10k: need transformations / analytical methods
  - >10k: need different approaches for "normal" and "high-scale"

Some keys: Categorical regions

- regions: contiguous bounded areas distinct from each other
  - using space to separate (proximity)
  - following expressiveness principle for categorical attributes

Idiom: bar chart

- one key, one value
  - colors
  - use order to separate (proximity)
  - eye: length = express count value
  - spatial regions: one per mark
    - separated boundarially aligned vertically
    - ordered by quartile
  - task
    - compare, looking values
    - scalability
    - dozen to hundreds of levels for key axis

Some keys: Express values (magnitudes)

0 Keys: Express values (magnitudes)

Idiom: scatterplot

- express values
  - quantitative attributes
  - no keys, only values
    - color
    - size (bubble plots)
  - channels
    - x-axis, y-axis position
    - size
    - number, sliders, distribution, correlation, clusters
    - scalability
    - hundreds of items

Scatterplots: Encoding more channels

- additional channels for point marks
  - color
  - size (bubble plots)
  - espaço: more axes grow quadratically radius is misleading
  - shape

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...

Notes:

- sorted: answer when done
- same scale?
- need different approaches based on scale
- homogeneity
- same/ different scales?
- need different approaches based on scale
- how many attributes?
- how many items?
- up to ~50: translatable with direct visual encoding
- thousands: need transformations / analytical methods
- up to ~10k: need different approaches for "normal" and "high-scale"

Dense Space-Filling
Separate Order Align
1 Key 2 Keys 3 Keys Many Keys
List Recursive Subdivision Volume Matrix
Rectilinear Parallel Radial
ARRANGE TABLES

AXIS ORIENTATION

Separate

Rectilinear

Idiom: heatmap

• two keys, one value
  – data
  – 2 case study (gene, experimental condition)
  – 1 quant attr (expression level)
  – marks: point
  – separate and align in 2D matrix
  – channel
    – color by quant attr
    – (default: diverging colormap)
  – task
    – find clusters, outliers
    – scalability
  – 194 terms, 100 of case levels, 10 quant attr 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-jointed by cluster hierarchy traversal
  – task: assess quality of cluster band by systematic methods

Idiom: pie chart, polar area chart

• pie chart
  – line mark with angle channel: variable (sector) width
  – separated & aligned radially, uniform 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

Idiom: glyphmaps

• rectilinear good for linear vs nonlinear trends

• radial good for cyclic patterns

Idiom: normalized stacked bar chart

• task
  – part-to-whole judgements

• normalized stacked bar chart
  – stacked bar chart, normalized to full size height
  – single stacked bar equivalent to full pie
  – high information density requires narrow rectangle

• pie chart
  – information density requires large rectangle

Idiom: parallel coordinates

• scatterplot matrix (SPLOM)
  – rectilinear axes, point mark
  – all possible pairs of axes
  – scalability
    – one dozen attributes
    – distance to hundreds of items

• space-filling idiom: heatmap
  • two keys, one value
  – data
  – 2 case study (gene, experimental condition)
  – 1 quant attr (expression level)
  – marks: point
  – separate and align in 2D matrix
  – channel
    – color by quant attr
    – (default: diverging colormap)
  – task
    – find clusters, outliers
    – scalability
  – 194 terms, 100 of case levels, 10 quant attr levels

Idiom: SPLOM

• scatterplot matrix
  – rectilinear axes, point mark
  – all possible pairs of axes
  – scalability
    – one dozen attributes
    – distance to hundreds of items

• space-filling idiom: heatmap
  • two keys, one value
  – data
  – 2 case study (gene, experimental condition)
  – 1 quant attr (expression level)
  – marks: point
  – separate and align in 2D matrix
  – channel
    – color by quant attr
    – (default: diverging colormap)
  – task
    – find clusters, outliers
    – scalability
  – 194 terms, 100 of case levels, 10 quant attr levels
**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

**Parallel coordinates quiz: car data**

- What correlations do you see?
  - positive?
  - negative?
  - none?
  - not sure?
  - horsepower to acceleration
  - weight to mileage?

**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
- parallel: unfamiliarity, training time

**Encode tables: Arrange space**

- data: text
  - text + 1 quant attrib per line
- derived data:
  - one pixel high
  - length according to original
  - color by attrib
  - scalability
  - 10K+ lines

**Idiom: Dense software overviews**

- Layout Density
  - Dense
  - Space-Filling

**Arrange tables**

- Express/Values
  - Separate, Order, Align-Regions
  - Parallel
  - Radial
  - Dense
  - Space-Filling

**Upcoming**

- D3 videos week 3
  - 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
- Foundations 3, out Thu Jan 30
- D3 videos/reading week 4
  - The General Update Pattern of D3.js (80 min)
  - Interaction with Unidirectional Data Flow (16 min)
  - Read/Reusable D3 Components

**Design critique & redesign: NZ**

- Consider the following questions:
  - What could be the goals of the designer for questions that the visualization answers? (domain-specific & abstract?)
  - What data is represented in this visualization? Be specific.
  - How is each data type visually encoded (morph/channels)?
  - Can you read the data precisely? Is the visual encoding appropriately chosen?
  - Will how would this work without numeric labels?
  - Develop two alternative designs to visualize this data.
    - free to discuss with your peers, but draw your own solution.
    - mark your best design, briefly note why you think it’s better.

**Credits**

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