### Information Visualization

#### Reduce: Aggregation & Filtering

**Project Peer Reviews**

**Tamara Munzner**  
Department of Computer Science  
University of British Columbia  

**Week 11, 17 Nov 2021**

**https://www.cs.ucl.ac.uk/~tamaras/347-31**

---

**Q&A / Backup Slides**

---

#### Visualization Analysis & Design

#### Reduce: Aggregation & Filtering (Ch 13)

**Tamara Munzner**  
Department of Computer Science  
University of British Columbia  
@tamaras

---

### How to handle complexity: 3 previous strategies + 1 more

#### Reduce: Aggregation & Filtering

<table>
<thead>
<tr>
<th>Device</th>
<th>Manipulate</th>
<th>Face</th>
<th>Justuxtapose</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Device Icon]</td>
<td>![Manipulate Icon]</td>
<td>![Face Icon]</td>
<td>![Justuxtapose Icon]</td>
<td>![Reduce Icon]</td>
</tr>
</tbody>
</table>

- **Derive new data to show within view**
- **Change view over time**
- **Facet across multiple views**
- **Reduce items/attributes within single view**

#### Filter

- **Eliminate some elements**
  - either items or attributes
  - according to what?
  - any possible function that partitions dataset into two sets
  - attribute values (larger/smaller than x)
  - menu/scroll

- **Filters vs queries**
  - query start with nothing, add in elements
  - filters: start with everything, remove elements
  - best approach depends on dataset size

#### Aggregate

- a group of elements is represented by a smaller number of derived elements

- **new table**: keys are bins, values are counts
- **opportunity for interaction**: control bin size on the fly

---

### How to handle complexity: 3 previous strategies

#### Reduce: Aggregation & Filtering

<table>
<thead>
<tr>
<th>Device</th>
<th>Manipulate</th>
<th>Face</th>
<th>Justuxtapose</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Device Icon]</td>
<td>![Manipulate Icon]</td>
<td>![Face Icon]</td>
<td>![Justuxtapose Icon]</td>
<td>![Reduce Icon]</td>
</tr>
</tbody>
</table>

- **Derive new data to show within view**
- **Change view over time**
- **Facet across multiple views**

### Idiom: FilmFinder

- **dynamic queries/filters for items**
  - tightly coupled interaction and visual encoding idioms, so user can immediately see results of action

### Idiom: cross filtering

- **team filtering**
- **coordinated views/controls combined**
  - all scented histogram/indicator updates when any ranges change

### Idiom: histogram

- **static item aggregation**
- **task/find distribution**
- **data table**
- **derived data**
  - new table lists are bins, values are counts
  - bin size
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly

### Idiom: scented widgets

- **augmented widgets show information scent**
  - better cues for information foraging whether value in drilling down further vs looking elsewhere
  - concise use of space: histogram on slider

---

**Upcoming**

- **next week (W12)**
  - async last week of readings/discussion (light, 2 readings)
  - Ch 14: Embed - Focus+Context
  - paper: Visualizing Deep Learning Models in TensorFlow
  - (type: design study)
  - in class post-update meetings with Tamara
  - oral feedback on project progress, after I've read them
  - last week of classes (W13)
    - async: last readings/discussion
      - in class: evals
      - in class: Q&A wrapup (W12)
    - in class: lecture on research process and final writeup expectations
Idiom: scented widgets
• augmented widgets show information scent
  – better cues for information foraging
  – show whether value in drilling down further vs. looking elsewhere
  – concise use of space: histogram on slider

Idiom: scented widgets
• augmented widgets show information scent
  – better cues for information foraging
  – show whether value in drilling down further vs. looking elsewhere
  – concise use of space: histogram on slider

Scented histogram bisliders: detailed

Idiom: boxplot
• static item aggregation
• task: find distribution
• data table
• derived data
  – 5 quant attributes
    • median: central line
    • lower and upper quartile boxes
    • lower and upper fences, whiskers
    • outliers beyond fence cutoffs explicitly shown
  – scalability: unlimited number of items

Figure 4: From left to right: box plot, vase plot, violin plot and bean plot. Within each plot, the distributions from left to right.

Dynamic aggregation: Clustering
• clustering: classification of items into similar bins
  – based on similarity measure
  – hierarchical algorithms produce "similarity trees": cluster hierarchy
  – aggregative clustering starts with each node as own cluster; then iteratively merge
  – cluster hierarchy: derived data used w/ many dynamic aggregation idioms
  – cluster more homogeneous than whole dataset
  – statistical measures & distribution more meaningful

Gerrymandering: MAUP for political gain
A real district in Pennsylvania: Democrats won 51% of the vote but only 5 out of 18 house seats

Dynamic aggregation: Spatial aggregation
• MAUP: Modifiable Areal Unit Problem
  – boundaries of cartographic regions can yield dramatically different results
  – zone effects
  – scale effects

Attribute aggregation: Dimensionality reduction
• attribute aggregation
  – derive low-dimensional target space from high-dimensional space
  – use when you can’t directly measure what you care about
  – true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
  – based on similarity measures, more meaningful

Gerrymandering: MAUP for political gain

Scented Widgets: Improving Navigation Cues with Information Scent

Idiom: Dimensionality reduction & visualization
• why do people do DR?
  – improve performance of downstream algorithm
  – avoid curse of dimensionality
  – data analysis
  – look in the output: visual data stream

Abstract tasks when visualizing DR data
• dimension-oriented tasks
  – naming synthesized dims, inspect data represented by loD points
• cluster-oriented tasks
  – verifying, naming, matching to classes
  – dynamic item aggregation
  – static item aggregation
  – cluster hierarchy

Dimension-oriented tasks
• naming synthesized dims: inspect data represented by lowD points
  – 5 quant attribs
  – median: central line
  – lower and upper quartile boxes
  – lower and upper fences, whiskers
  – outliers beyond fence cutoffs explicitly shown
  – scalability: unlimited number of items

Cluster-oriented tasks
• verifying, naming, matching to classes
  – dynamic item aggregation
  – static item aggregation
  – cluster hierarchy

Dynamic aggregation: Dimensioning
• vocab use in field not consistent
– dimension/attribute
• attribute...
VDA with DR example: nonlinear vs linear

- **DR for computer graphics reflectance model**
  - goal: simulate how light bounces off materials to make realistic pictures
  - ideas: measure what light does with real materials
  - many techniques proposed
  - note: dim estimate depends on technique used!
  - second try: charting (nonlinear DR technique)
    - scree plot suggests 10-15 dims
    - note: dim estimate depends on technique used!

Capturing & using material reflectance

- reflectance measurement: interaction of light with real materials (spheres)
- results: 104 high-res images of material
- each image 4M pixels
- goal: image synthesis
  - simple: create new materials
  - need for more concise model
    - 104 materials * 4M pixels = 400M dims
    - want concise model with meaningful knobs
    - how dirty/greasy/metallic?
    - DR to the rescue!

Linear DR

- first try: PCA (linear)
  - result: error falls off sharply after ~45 dimensions
  - problem: physically impossible intermediate points when simulating new materials
  - specular highlights cannot have holes!

Nonlinear DR

- second try: charting (nonlinear DR technique)
  - scree plot suggests 10-15 dims
  - note: dim estimate depends on technique used!