Week 5: Manipulate, Facet, Reduce
Demo: Text

Tamara Munzner
Department of Computer Science
University of British Columbia

JRNL 520M, Special Topics in Contemporary Journalism: Visualization for Journalists
Week 5: 13 October 2015

http://www.cs.ubc.ca/~tmm/courses/journ15
Now

• Manipulate
• Facet (not covered last week)
• Reduce
• Demos/Videos
  – LineUp
  – LiveRAC
  – Cerebral
• Demos: Text
  – Overview
  – TimeLineCurator
**Encode**

- **Arrange**
  - Express
  - Separate
- **Order**
  - Align
- **Use**

**Map**
from categorical and ordered attributes

- **Color**
  - Hue
  - Saturation
  - Luminance
- **Size, Angle, Curvature, ...**
- **Shape**
  - □ □ □ □
- **Motion**
  - Direction, Rate, Frequency, ...

**Manipulate**

- **Change**

**Facet**

- **Juxtapose**

**Reduce**

- **Filter**
- **Aggregate**
- **Embed**
How to handle complexity: 3 more strategies + 1 previous

- **Manipulate**
  - Change
  - Select
  - Navigate

- **Facet**
  - Juxtapose
  - Partition
  - Superimpose

- **Reduce**
  - Filter
  - Aggregate
  - Embed

- **Derive**

- • change view over time
- • facet across multiple views
- • reduce items/attributes within single view
- • derive new data to show within view
How to handle complexity: 3 more strategies  + 1 previous

**Manipulate**

- **Change**
  - ![Change Diagram]

- **Select**
  - ![Select Diagram]

- **Navigate**
  - ![Navigate Diagram]

**Facet**

- **Juxtapose**
  - ![Juxtapose Diagram]

- **Partition**
  - ![Partition Diagram]

- **Superimpose**
  - ![Superimpose Diagram]

**Reduce**

- **Filter**
  - ![Filter Diagram]

- **Aggregate**
  - ![Aggregate Diagram]

- **Embed**
  - ![Embed Diagram]

**Derive**

- ![Derive Diagram]

- **change over time**
  - most obvious & flexible of the 4 strategies
VAD Ch 11: Manipulate

- **Change over Time**

- **Navigate**
  - **Item Reduction**
    - **Zoom**
      - Geometric or Semantic
  - **Pan/Translate**
  - **Constrained**

- **Select**
Change over time

• change any of the other choices
  – encoding itself
  – parameters
  – arrange: rearrange, reorder
  – aggregation level, what is filtered...

  – interaction entails change
Idiom: Re-encode

System: Tableau

made using Tableau, http://tableausoftware.com
Idiom: **Reorder**

- data: tables with many attributes
- task: compare rankings

System: **LineUp**

Idiom: **Realign**

- stacked bars
  - easy to compare
    - first segment
    - total bar
- align to different segment
  - supports flexible comparison

System: **LineUp**

**Idiom: Animated transitions**

- smooth transition from one state to another
  - alternative to jump cuts
  - support for item tracking when amount of change is limited
- example: multilevel matrix views
  - scope of what is shown narrows down
    - middle block stretches to fill space, additional structure appears within
    - other blocks squish down to increasingly aggregated representations

Navigate: Changing item visibility

• change viewpoint
  – changes which items are visible within view
  – camera metaphor
    • zoom
      – geometric zoom: familiar semantics
      – semantic zoom: adapt object representation based on available pixels
        » dramatic change, or more subtle one
    • pan/translate
    • rotate
      – especially in 3D

– constrained navigation
  • often with animated transitions
  • often based on selection set
Idiom: **Semantic zooming**

- visual encoding change
  - colored box
  - sparkline
  - simple line chart
  - full chart: axes and tickmarks

**System: LiveRAC**

VAD Chap 11: Facet Into Multiple Views

- Juxtapose

- Partition

- Superimpose
How to handle complexity: 3 more strategies

Manipulate

- Change
- Select
- Navigate

Facet

- Juxtapose
- Partition
- Superimpose

Reduce

- Filter
- Aggregate
- Embed

Derive

- facet data across multiple views
Facet

- **Juxtapose**
  - ![Juxtapose Diagram]

- **Partition**
  - ![Partition Diagram]

- **Superimpose**
  - ![Superimpose Diagram]

Coordinate Multiple Side By Side Views

- **Share Encoding: Same/Different**
  - ![Share Encoding Diagram]

- **Share Data: All/Subset/None**
  - ![Share Data Diagram]

- **Share Navigation**
  - ![Share Navigation Diagram]
Idiom: **Linked highlighting**

- see how regions contiguous in one view are distributed within another
  - powerful and pervasive interaction idiom

- encoding: different
  - **multiform**

- data: all shared

Idiom: bird’s-eye maps

- encoding: same
- data: subset shared
- navigation: shared
  - bidirectional linking

- differences
  - viewpoint
  - (size)

- overview-detail

System: Google Maps

Idiom: **Small multiples**

- **encoding:** same
- **data:** none shared
  - different attributes for node colors
  - (same network layout)
- **navigation:** shared

[**System: Cerebral**](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2579090/)

Coordinate views: Design choice interaction

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Subset</td>
<td>None</td>
</tr>
<tr>
<td><strong>Encoding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>Redundant</td>
<td>Overview/Detail</td>
<td>Small Multiples</td>
</tr>
<tr>
<td>Different</td>
<td>Multiform</td>
<td>Multiform, Overview/Detail</td>
<td>No Linkage</td>
</tr>
</tbody>
</table>

• why juxtapose views?
  – benefits: eyes vs memory
    • lower cognitive load to move eyes between 2 views than remembering previous state with single changing view
  – costs: display area, 2 views side by side each have only half the area of one view
Partition into views

• how to divide data between views
  – encodes association between items using spatial proximity
  – major implications for what patterns are visible
  – split according to attributes

• design choices
  – how many splits
    • all the way down: one mark per region?
    • stop earlier, for more complex structure within region?
  – order in which attribs used to split
  – how many views
Partitioning: List alignment

• single bar chart with grouped bars
  – split by state into regions
    • complex glyph within each region showing all ages
  – compare: easy within state, hard across ages

• small-multiple bar charts
  – split by age into regions
    • one chart per region
  – compare: easy within age, harder across states
Partitioning: Recursive subdivision

- split by neighborhood
- then by type
- then time
  - years as rows
  - months as columns
- color by price

- neighborhood patterns
  - where it’s expensive
  - where you pay much more for detached type

System: HIVE

Partitioning: Recursive subdivision

• switch order of splits
  – type then neighborhood

• switch color
  – by price variation

• type patterns
  – within specific type, which neighborhoods inconsistent

Partitioning: Recursive subdivision

- different encoding for second-level regions
  - choropleth maps

Superimpose layers

• **layer**: set of objects spread out over region
  – each set is visually distinguishable group
  – extent: whole view

• design choices
  – how many layers?
  – how are layers distinguished?
  – small static set or dynamic from many possible?
  – how partitioned?
    • heavyweight with attribs vs lightweight with selection

• distinguishable layers
  – encode with different, nonoverlapping channels
    • two layers achieveable, three with careful design
Static visual layering

- foreground layer: roads
  - hue, size distinguishing main from minor
  - high luminance contrast from background
- background layer: regions
  - desaturated colors for water, parks, land areas
- user can selectively focus attention
- “get it right in black and white”
  - check luminance contrast with greyscale view

Superimposing limits

• few layers, but many lines
  – up to a few dozen
  – but not hundreds

• superimpose vs juxtapose: empirical study
  – superimposed for local visual, multiple for global
  – same screen space for all multiples, single superimposed
  – tasks
    • local: maximum, global: slope, discrimination

Dynamic visual layering

- interactive, from selection
  - lightweight: click
  - very lightweight: hover

- ex: 1-hop neighbors

Reduce items and attributes

- reduce/increase: inverses
- filter
  - pro: straightforward and intuitive
    - to understand and compute
  - con: out of sight, out of mind
- aggregation
  - pro: inform about whole set
  - con: difficult to avoid losing signal
- not mutually exclusive
  - combine filter, aggregate
  - combine reduce, change, facet

Reducing Items and Attributes

- Filter
  - Items
  - Attributes
- Aggregate
  - Items
  - Attributes
- Embed
Idiom: **dynamic filtering**

- item filtering
- browse through tightly coupled interaction
  - alternative to queries that might return far too many or too few

Idiom: **histogram**

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

- static item aggregation
- data: table
- derived data: table
  - key attribs x,y for pixels
  - quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential
categorical hue + ordered luminance colormap

Idiom: boxplot

- static item aggregation
- task: find distribution
- data: table
- derived data
  - 5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - values beyond which items are outliers
  - outliers beyond fence cutoffs explicitly shown

[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]
Idiom: **Hierarchical parallel coordinates**

- dynamic item aggregation
- derived data: *hierarchical clustering*
- encoding:
  - cluster band with variable transparency, line at mean, width by min/max values
  - color by proximity in hierarchy

Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - gerrymandering (manipulating voting district boundaries) is one example!

[http://www.e-education.psu.edu/geog486/l4_p7.html, Fig 4.cg.6]
Dimensionality reduction

• attribute aggregation
  – derive low-dimensional target space from high-dimensional measured space
  – use when you can’t directly measure what you care about
  • true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
  • latent factors, hidden variables

Tumor Measurement Data ➔ DR ➔ Benign

data: 9D measured space

derived data: 2D target space
Dimensionality reduction for documents

Task 1
In HD data → Out 2D data

What? In High-dimensional data
Why? Produce Derive

Task 2
In 2D data → Out Scatterplot Clusters & points

What? In 2D data
Why? Discover Explore
How? Encode Navigate Identify Select

Task 3
In Scatterplot Clusters & points → Out Labels for clusters

What? In Scatterplot Clusters & points
Why? Produce Annotate

• bag of words model for text document
Overview origin story: WikiLeaks meets Glimmer

• WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
  – conjecture that existing label classification falls short of showing all meaningful structure in data
    • friendly action, criminal incident, ...
  – had some NLP, needed better vis tools

• Glimmer: multilevel dimensionality reduction algorithm
  – scalability to 30K documents and terms

Overview design evolution

• how to find the needle in the haystack?
• how to convince that the haystack has no needles?
What/Why/How interplay

- why: understand clusters
- what: derive data of full cluster hierarchy
  - explore space of possible clusterings
- how: show cluster hierarchy
  - arrange space: node-link
- how: support tagging clusters/docs
  - following or cross-cutting hierarchy!
    - simple annotation
    - progress tracking
    - user-defined semantics

Dataset Types

- Networks
- Trees

Targets

- Network Data
  - Topology
  - Paths

Produce

Annotate

Arrange Networks And Trees
How: Idiom design decisions

- facet: juxtapose linked views
  - linked color coding
    - cluster hierarchy tree
    - DR scatterplot
    - tags
  - reading text/keywords
    - cluster list
    - doc reader

- Juxtapose and Coordinate Views
  - Share Encoding: Same/Different
  - Share Data: All/Subset/None
  - Linked Highlighting

Why?

What?

• facet: juxtapose linked views
  - linked color coding
    • cluster hierarchy tree
    • DR scatterplot
    • tags
  - reading text/keywords
    • cluster list
    • doc reader

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape
Overview video (version 1)

http://www.cs.ubc.ca/labs/imager/tr/2012/modiscotag/
Overview video v4

- versions 3 and 4
  - no DR scatterplot
  - tree arrangement emphasizing nodes not links
  - combined doc/cluster viewer

http://vimeo.com/71483614
Why: Task abstractions

• what’s in this collection? (of leaked docs)
  – generate hypothesis
  – summarize clusters
  – explore clusters

• locate evidence (within FOIA dump)
  – verify hypothesis
  – identify clusters/documents
  – locate clusters/documents

• prove non-existence of evidence
  – even harder!
  – exhaustive reading vs filtering out irrelevant

http://www.cs.ubc.ca/labs/imager/tr/2014/Overview/
Further reading

  – Chap 11: Manipulate View
  – Chap 12: Facet Across Multiple Views
  – Chap 13: Reduce Items and Attributes
Lab/Assignment 5

• Use TimeLineCurator to create visual timelines from free-form text
  – work through BC History example
  – find 1 article where temporal story is worth telling, and curate it for TimelineJS export
    ● including media/images is optional
  – find 2 articles that make sense to compare with each other in a mashup
    ● curate a combined timeline for TLC export
  – find 1 article where there’s nothing interesting to see
    ● document that it’s uninteresting with screenshot of TLC’s initial screen