Data and Dataset Types

Data Types

Tables

Attributes (columns)

Trees

Attributes

Positions

Geometry

What?

Attributes (columns)

Value in cell

Clusters,

Diverging

Quantitative

Ordinal

Internal or External Representation

External Representation: Replace Cognition with Perception

Why?

Computer-based visualization systems designed to help people carry out tasks more effectively.

What?

Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

Why represent all the data?

Contents

Why focus on tasks and effectiveness?

Why analyze?

Why focus on tasks and effectiveness?

Why analyze?

• summaries lose information, details matter
• confirm expected and find unexpected patterns
• assess validity of statistical model

Anscocomb's Quartet

Identical statistics


Further reading

• Ch9. Why focus on tasks and effectiveness?
• Why represent all the data?
• Why focus on tasks and effectiveness?
• Why analyze?
• Why focus on tasks and effectiveness?
• Why analyze?
• Why focus on tasks and effectiveness?
• Why analyze?
### Types: Datasets and data types

- **Dataset Types**
  - Tables
  - Networks
- **Attributes**
  - Categorical
  - Quantitative

#### Three major data types
- **Tables**
  - Fields
  - Geometry
- **Networks**
  - Fields (continuous)
  - Geometry (spatial)
- **Spatial**
  - Fields (continuous)
  - Geometry (spatial)

#### Data and dataset types
- **Data and Dataset Types**
  - Tables
  - Networks & Trees
  - Fields
  - Geometry
  - Clusters, Sets, Lists

#### Attributes
- **Attributes**
  - Links
  - Positions
  - Grids

### Actions: Analyze, Query

- **Consume**
  - discover vs present
  - alias explore vs explain
- **Analyze**
  - discover
  - annotate, record
  - derive
- **Produce**
  - quantitate
  - analyze
  - query

### Attribute types
- **Categorical**
  - Ordinal
  - Quantitative

### Ordering Direction
- **Sequential**
- **Diverging**
- **Cyclic**

### Further reading, full Ch 2
  - The Eyes Have It: A Taxonomy for Information Visualization, Ben Shneiderman. Proc. IEEE InfoVis 1996.

### Further reading, full Ch 3
Discriminability: How many usable steps?
- must be sufficient for number of attribute levels to show
  - linewidth: few bns but salient
  - expression principle
  - match channel and data characteristics
  - expressiveness principle
  - encode most important attributes with highest ranked channels

Separability vs. Integrity
- perceptual system mostly operates with relative judgements, not absolute
  - Weber's Law: ratios of increment to background is constant
  - filled rectangles differ in length by 1.2, easy judgement

Relative vs. absolute judgements
- perception of luminance is contextual based on contrast with surroundings
- speed independent of distractor count
- speed depends on channel and amount of difference from distractors
- serial search for (almost all) combinations
- speed depends on number of distractors

Further Reading

Ch 6. Rules of Thumb
- No unjustified 3D
  - Power of the plane
  - Disparity of depth
  - Occlusion hides information
  - Perspective distortion dangers
  - Tilted text isn't illegible
- No unjustified 2D
  - Eyes beat memory
  - Resolution over immersion
  - Overview first, zoom and filter, details on demand
  - Responsiveness is required
  - Function first, form next

Further Reading
Perspective distortion loses information

- perspective distortion
  - interferes with all size channel encodings
  - power of the plane is lost!

3D vs 2D bar charts

- 3D bars very difficult to justify!
  - perspective distortion
  - occlusion
  - faceting into 2D almost always better choice

Justified 3D: shape perception

- benefits outweigh costs when task is shape perception for 3D spatial data
  - interactive navigation supports synthesis across many viewpoints

No unjustified 3D example: Time-series data

- extruded curves: detailed comparisons impossible

Justified 3D: Economic growth curve

- constrained navigation steps through carefully designed viewpoints

Eyes beat memory

- principle: external cognition vs. internal memory
  - easy to compare by moving eyes between side-by-side views
  - harder to compare visible item to memory of what you saw

Implications for animation

- great for choreographed storytelling
  - great for transitions between states
- poor for many states with changes everywhere
  - consider small multiples instead

Eyes beat memory example: Cerebral

- small multiples: one graph instance per experimental condition
  - same spatial layout
  - color differently by condition

Why not animation?

- disparate frames and regions: comparison difficult
  - vs contiguous frames
  - vs small region
  - vs coherent motion of group

 safe special case
- animated transitions
Arrange Tables
Separate, Order, Align Regions
Express Values
Axis Orientation
Layout Density
Express Values
Rectilinear
Parallel
Radial

Overview first, zoom and filter, details on demand
• influential mantra from Shneiderman
  • overview = summary
  • microcosm of full vis design problem
  • layout density
  • dense space-filling
  • separation order align
  • 1 key 2 keys 3 keys many keys
  • list recursive subdivision
  • volume matrix
  • rectilinear parallel radial

Rule of thumb: Responsiveness is required
• 0.1 seconds: perceptual processing
• 0.2 seconds: response for meaningful highlighting - ballistic reaction
• 0.5 seconds: immediate response
• fast response after mouseclick, button press - Fitt's Law limits on mouse control
• 1 second: brief task
• bounded response after dialog box - mental model of heavyweight operation (file load)
• scalability considerations
  • highlight selection with complete redraw of view (graphics framebuffer)
  • show houghs for multi-second operations (check for cancel/undo)
  • show progress bar for long operations (process in background thread)
  • rendering speed when item count is large (guaranteed frame rate)

Function first, form next
• start with focus on functionality
  • possible to improve aesthetics later on, as refinement
  • if no expertise in-house, find good graphic designer to work with
  • aesthetics do matter: another level of function
  • visual hierarchy, alignment, flow
    • Gestalt principles in action
  • dangerous to start with aesthetics
    • usually impossible to add function retroactively

Form: Basic graphic design principles
• projecting
  • do group-related items together
  • avoid equal whitespace between unrelated items
• algorithm
  • do find shortest strong line, stick to it
  • avoid automatic centering
• repetition
  • do unify by pushing existing consistencies
• contrast
  • if few identical, then very different
  • avoid similar
• layout density
  • buy now and read cover to cover - very practical, worth your time, fast read!

Best practices: Labelling
• make visualizations as self-documenting as possible
• meaningful & useful title, labels, legends
• axes and panels/axes should have labels
• and axes should have good readable boundary tick marks
• everything that's plotted should have a legend
• and over-labeling is not redundant with main title
• use reasonable number of fonts
• avoid scientific notation in most cases

Further reading
• Visualization and Analysis and Design. Tamara Munzner. CRC Press, 2014. - Chap 4: Rules of Thumb

Encode tables: Arrange space
• encode
  • express
  • separate
  • order
  • align
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key
  • key

Ch 7. Arrange Tables
Idiom: bar chart
- one key, one value
  - data
  - 1 categ attrib, 1 quant attrib
  - mark: line
  - channels
    - length: express quant value
    - spatial regions: one per mark
  - separated horizontally aligned vertically
    - ordered by quant attribs
      - by label (alphabetically), by length attrib (data-driven)
  - task
    - compare, lookup values
    - scalability
  - dozens to hundreds of keys for each mark

LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]
**Idiom: similarity-clustered streamlines**

- data
  - 3D vector field
- derived data (from field)
  - similarity—streamlines will follow
- derived data (per streamline)
  - curvature, torsion, torsority
- agnostic: complex, weighted combination
- compute cluster hierarchy across all signatures
- encode color and opacity by cluster
- tasks
  - find features, query shape
- scalability
  - millions of samples, hundreds of streamlines

**Further reading**

- Chap 8: Arrange Speed Dots
- Further reading, full

**Idiom: force-directed placement**

- visual encoding
  - link connection marks, node point marks
- considerations
  - special position on meaning directly encoded
  - left to too extensive zoom
  - proximity semantics
  - variations of other tools
  - various tabular, artful lab of layout algorithms
  - weighs and flex
  - long edges carry more than short
- tasks
  - explore topology, locate paths, clusters
- scalability
  - node/edge density < 40

**Further reading**

- Data: network
  - derived cluster hierarchy
  - better algorithm for some encoding technique
  - some fundamental use of space
  - hierarchy used for algorithm sensitivity that is not shown explicitly
  - use an algorithm vs encoding in itself
- scalability
  - nodes, edges: IK-10K
  - hard problem eventually hits

**Idiom: radial node-link tree**

- data
  - tree
- encoding
  - link connection marks
  - point node marks
  - radial arc orientation
  - angular proximity vector
- tasks
  - understand topology, following path
- scalability
  - IK-10K nodes

**Further reading**

- Data: network
  - 1 quant attr
  - weight edge between nodes
  - 2-cang attrb: node less x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - IK nodes, IAM edges

**Idiom: sfdp (multi-level force-directed placement)**

- data: network
- transform into same data encoding as heapmap
- derived data: table from network
  - 1 quant attrb
  - weighted edge between nodes
  - 2-cang attrb: node less x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - IK nodes, IAM edges

**Idiom: adjacency matrix view**

- marks as links (vs notes)
  - common case in network drawing
- ID case: connection
  - all node-link diagrams
- emphasize topology path tracing
  - networks and trees
- 2D case: containment
  - all tree views
  - emphasize attribute values at leaves (node size)
  - only trees

**Link marks: Connection and containment**

- Ch 10. Map Color and Other Channels

**Tree drawing idioms comparison**

- data shown
  - link relationship
  - tree depth
  - sibling order
- design choices
  - controls vs containment/link marks
  - rectilinear vs radial layout
- spatial position choices
- considerations
  - redundant? arbitrary?
- information density?
- avoid wasting space

**Further reading**

- Chap 8: Arrange Speed Dots
- Further reading, full

**Ch 9. Arrange Networks and Trees**

- Node-Link Diagrams
  - node matrix
  - adjacency matrix
  - containment
  - tree matrix
- Enclosure
  - containment
- node-link diagram strengths
  - derived data: table from network
  - visual encoding
  - scalability
  - node/edge density < 40
- connection vs. adjacency comparison
  - predictability, scalability, supports reordering
  - some topology tasks trainable
  - node-link diagram strengths
  - topology understanding, path tracing
  - intuitive, no training needed
  - empirical study
  - node-link best for small networks
  - matrix best for large networks

- If tasks don’t involve topological structure!
  - the readability of graphs using scalable and topological representations, a conceded experiment using statistical analysis.


**Connection vs. Adjacency vs. Similarity**

- see discussion in earlier papers
- this chapter is more of a summary and overview of the principles

- Figure 7.5 shows a simple example of an undirected network, with a relation from node A to node B necessarily implies a link from B to A. For directed networks, an associated quantitative value attribute, by encoding with an ordered channel metric, a value.

Spectral sensitivity

Luminance
- Need luminance for edge detection
  - Fine-grained detail only visible through luminance contrast
  - Legible text requires luminance contrast!
- Intrinsic perceptual ordering

Opponent color and color deficiency
- Perceptual processing before optic nerve
  - One achromatic luminance channel (L)
  - Edge detection through luminance contrast
  - 2 chromatic channels
    - Red-green (r/g) and yellow-blue (b/y)
    - “Color blind”: one axis has degraded acuity
      - 8% of men are red/green color deficient
- Color deficiency: Reduces color to 2 dimensions

Color spaces
- CIELUV good for computation
  - L*: intensity, perceptually linear luminance
  - a*: intensity, perceptually linear but non-monotonic
  - b*: intensity, poor for encoding
- HLS/HSV somewhat better for encoding
  - Hue: only pseudo-perceptual
  - Lightness (L) or value (V) + luminance or L*
- Luminance, hue, saturation
  - Good for encoding
  - But not standard graphics tools colorspace

Designing for color deficiency: Check with simulator
Normal
Deuteranope
Protanope
Tritanope

http://rehue.net

Designing for color deficiency: Avoid encoding by hue alone
- Redundantly encode
  - Vary luminance
  - Change shape

Relative judgements: Color & illumination
- Do they match?
- Oh no, that looks wrong!

Categorical color: Limited number of discriminable bins
- Human perception built on relative comparisons
  - Great if color contiguous
  - Surprisingly bad for absolute comparisons
- Noncontiguous small regions of color
  - Fewer bins than you want
  - Rule of thumb: 6-12 bins, including background and highlights

Color Encoding
- Encode
  - Manipulate
  - How?

Image courtesy of John McCann

Categorical vs ordered color
- Order can show magnitude
  - Luminance: how bright
  - Saturation: how colorful
- Categorical can show identity
  - Hue what color

Channels have different properties
- What they convey directly to perceptual system
  - How much they can convey: how many discriminable bins can we use?

Bezold Effect: Outlines matter
- Color constancy: simultaneous contrast effect

Relative judgements: Color & illumination
- Good vs. different
  - Image courtesy of John McCann

Decomposing color
- First rule of color: do not talk about color!
  - Color is confusing if treated as monolithic
- Decompose into three channels
  - Order can show magnitude
  - Luminance: how bright
  - Saturation: how colorful
  - Categorical can show identity
  - Hue, what color
- Channels have different properties
  - What they convey directly to perceptual system
  - How much they can convey: how many discriminable bins can we use?
Further reading, full


Further reading

**Idiom: Semantic zooming**
- **semantic zoom**
  - alternative to geometric zoom
  - resolution-aware layout adapts to available space
  - goal legibility at multiple scales
  - dramatic or subtle effects
  - visual encoding change
    - colored box
    - sparkline
    - simple line chart
    - full chart axes and tick marks

**System: LiveRAC**
- **semantic zoom**
  - alternative to geometric zoom
  - resolution-aware layout adapts to available space
  - goal legibility at multiple scales
  - dramatic or subtle effects
  - visual encoding change
    - colored box
    - sparkline
    - simple line chart
    - full chart axes and tick marks

**Navigate: Reducing attributes**
- **continuation of camera metaphor**
  - show only items matching specific value for given attribute alone
  - axis signal or inter-axis alignment
  - cut
  - show only items for side of plane from camera
  - project
  - show mathematics of image creation
    - orthographic (principal 3rd dimension)
    - perspective (transmogrifying camera line 3D information)

**Rule of thumb: Responsiveness is required**
- **visual feedback three rough categories**
  - 0-1 seconds: perceptual processing
  - 1 second: brief tasks
  - 1-2 seconds: simple tasks

**Manipulate**
- **Navigate**
  - Item Reduction
  - Attribute Reduction
  - Parent/Children
  - Full/Partial

**Idiom: Semantic zooming**
- **semantic zoom**
  - alternative to geometric zoom
  - resolution-aware layout adapts to available space
  - goal legibility at multiple scales
  - dramatic or subtle effects
  - visual encoding change
    - colored box
    - sparkline
    - simple line chart
    - full chart axes and tick marks

**Navigate: Changing viewpoint/visibility**
- **change viewpoint**
  - changes which items are visible within view
    - full-screen
    - pan/translate/scroll
    - move up/down/sideways

**Idiom: Scrollytelling**
- **how**: navigate page by scrolling (panning down)
- **why**: changes which items are visible within view
  - familiar & intuitive, from standard web browsing
  - linear (up & down) vs possible overload of click-based interface choices

**Idiom: Animated transition + constrained navigation**
- **example**: geographic map
  - simple zoom, only viewport changes, shapes preserved

**Interaction benefits**
- **interaction pros**
  - major advantage of computer-based vs paper-based visualization
  - visual feedback three rough categories
    - 0-1 seconds: perceptual processing
    - 1 second: brief tasks
    - 1-2 seconds: simple tasks

**Navigate: Unconstrained vs constrained**
- **unconstrained navigation**
  - easy to implement for designer
  - hard to control for user
  - easy to overshoot/undershoot
  - typically uses animated transitions

**Idiom: Animated transition + constrained navigation**
- **example**: icicle plot
  - transition into containing mark causes aspect ratio (shape) change

**Navigate: Changing viewpoint/visibility**
- **change viewpoint**
  - changes which items are visible within view
  - full-screen
  - pan/translate/scroll
  - move up/down/sideways
  - rotate/spin
  - typically in 3D
  - zoom in/out
  - perspective (foreshortening captures limited 3D information)
  - unusual channels: motion
    - visual encoding change
    - fluid task switching: different visual encodings support different tasks
    - contiguity: no interruptions
    - responsiveness is required
    - can provide useful additional detail on demand
    - render speed when item count is large (guaranteed frame rate)

**Interaction benefits**
- **interaction pros**
  - major advantage of computer-based vs paper-based visualization
  - visual feedback three rough categories
    - 0-1 seconds: perceptual processing
    - 1 second: brief tasks
    - 1-2 seconds: simple tasks
  - fluid task switching: different visual encodings support different tasks
  - animated transitions provide excellent support

**Idiom: Scrollytelling**
- **how**: navigate page by scrolling (panning down)
- **why**: changes which items are visible within view
  - familiar & intuitive, from standard web browsing
  - linear (up & down) vs possible overload of click-based interface choices

**Navigate: Cartographic projections**
- **project**: from 3D sphere surface to 2D plane
  - can only fully preserve 2 out of 3
    - angles conformal
    - area equal area
    - conformality/areas
  - added detail during transition
  - typically in 3D
  - zoom in/out
  - perspective (transmogrifying camera line 3D information)

**Idiom: Animated transition + constrained navigation**
- **example**: multilevel matrix views
  - add detail during transition
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Zoomin.avi
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Zoomout.avi
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Pan.avi

**Idiom: Animated transition + constrained navigation**
- **example**: multilevel matrix views
  - add detail during transition
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Zoomin.avi
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Zoomout.avi
  - movie: http://www.win.tue.nl/vlvi/home/hans/matrix/Pan.avi

**Navigate: Changing viewpoint/visibility**
- **change viewpoint**
  - changes which items are visible within view
  - full-screen
  - pan/translate/scroll
  - move up/down/sideways
  - rotate/spin
  - typically in 3D
  - zoom in/out
  - perspective (foreshortening captures limited 3D information)
  - unusual channels: motion
    - visual encoding change
    - fluid task switching: different visual encodings support different tasks
    - contiguity: no interruptions
    - responsiveness is required
    - can provide useful additional detail on demand
    - render speed when item count is large (guaranteed frame rate)

**Navigate: Cartographic projections**
- **project**: from 3D sphere surface to 2D plane
  - can only fully preserve 2 out of 3
    - angles conformal
    - area equal area
    - conformality/areas
  - added detail during transition
  - typically in 3D
  - zoom in/out
  - perspective (transmogrifying camera line 3D information)

**Navigate: Scrollytelling**
- **how**: navigate page by scrolling (panning down)
- **why**: changes which items are visible within view
  - familiar & intuitive, from standard web browsing
  - linear (up & down) vs possible overload of click-based interface choices

**Idiom: Semantic zooming**
- **semantic zoom**
  - alternative to geometric zoom
  - resolution-aware layout adapts to available space
  - goal legibility at multiple scales
  - dramatic or subtle effects
  - visual encoding change
    - colored box
    - sparkline
    - simple line chart
    - full chart axes and tick marks

**Idiom: Animated transition + constrained navigation**
- **example**: geographic map
  - simple zoom, only viewport changes, shapes preserved

**Navigate: Reducing attributes**
- **continuation of camera metaphor**
  - show only items matching specific value for given attribute alone
  - axis signal or inter-axis alignment
  - cut
  - show only items for side of plane from camera
  - project
  - show mathematics of image creation
    - orthographic (principal 3rd dimension)
    - perspective (transmogrifying camera line 3D information)

**Rule of thumb: Responsiveness is required**
- **visual feedback three rough categories**
  - 0-1 seconds: perceptual processing
  - 1 second: brief tasks
  - 1-2 seconds: simple tasks

**Manipulate**
- **Navigate**
  - Item Reduction
  - Attribute Reduction
  - Parent/Children
  - Full/Partial

**Idiom: Scrollytelling**
- **how**: navigate page by scrolling (panning down)
- **why**: changes which items are visible within view
  - familiar & intuitive, from standard web browsing
  - linear (up & down) vs possible overload of click-based interface choices

**Navigate: Cartographic projections**
- **project**: from 3D sphere surface to 2D plane
  - can only fully preserve 2 out of 3
    - angles conformal
    - area equal area
    - conformality/areas
  - added detail during transition
  - typically in 3D
  - zoom in/out
  - perspective (transmogrifying camera line 3D information)
Facet

Further reading

- MicroView.

Further reading, full


Linked views

- unidirectional vs bidirectional linking

Linked views: Multidirectional linking

- System: Buckets

Linked views: Multiform views

- System: PathFinder

Complex linked multiform views

- System: StratomeX

Flows: R/Shiny

Idiom: Overview-detail navigation

- encoding: same
- data: subset shared
- navigation: shared
- unidirectional linking
- special case: birds-eye map

Idiom: Overview-detail views

- encoding: same
- data: subset shared
- navigation: shared
- bidirectional linking
- special case: birds-eye map

Video: Visual Analysis of Historical Hotel Visitation Patterns

- System: Miscrosoft

Overview-detail

- multiscle: three viewing levels
- linked views
- dynamic filtering
- tooling processing (modern version p5.js.org)

Overview-detail

- unidirectional linking

Overview-detail

- unidirectional linking

Overview-detail

- unidirectional linking

Features

- linked data
- dynamic filtering
- tooling processing
- (modern version p5.js.org)
Reduce items and attributes

- **Reduce/increase:** inverses
- **Filter**
  - pros: straightforward and intuitive
  - cons: difficult to avoid losing signal
- **Aggregate**
  - combine filter, aggregate
  - reduce, change, facet

**Idiom:** cross filtering

- **Item filtering**
- **Coordinated views/controls combined**
- all-scanned histogram builders update when any ranges change

**Idiom:** histogram

- **Static item aggregation**
- **Task find distribution**
- **Data table**
- **Derived data**
  - new table: keys are bins, values are counts
  - bin size crucial
- **Pattern change dramatically depending on discretization**
- opportunity for interaction control bin size on the fly

**Spatial aggregation**

- **MAUP:** Modifiable Areal Problem
  - Gerrymandering (manipulating voting district boundaries) is only one example!
- some effects

**Dimensionality reduction**

- **Attribute aggregation**
  - derive low-dimensional target space from high-dimensional measured data
  - capture most of variance with minimal error

**Idiom:** hierarchical parallel coordinates

- **Dynamic item aggregation**
- **Task find distribution**
- **Data table**
- **Derived data**
- cluster with variable transparency line at mean, width by min/max values
- **Color by proximity in hierarchy**

**Idiom:** aggregation via hierarchical clustering (visible)

- **Attribute aggregation**
  - derive low-dimensional target space from high-dimensional measured space
  - capture most of variance with minimal error
  - use when you can't directly measure what you care about
  - random dimensionality of data compressed to be closer dimensionality of measurements
  - latent factors, hidden variables

**Idiom:** boxplot

- **Static item aggregation**
- **Task find distribution**
- **Data table**
- **Derived data**
  - 5-quantile attributes
  - medians, control line
  - lower and upper quartile boxes
  - lower and upper fences/whiskers

- values beyond these are outliers
- outliers beyond fence cutoffs explicitly shown

**Idiom:** smented widgets

- augmented widgets show information scent
- cues to show whether value in drilling down further or looking elsewhere
- **Concise use of space:** histogram on slider
Dimensionality vs attribute reduction

- vocab use in field is not consistent
  - dimension/attribute
  - attribute reduction: reduce set with filtering
  - includes orthographic projection
- dimensionality reduction: create smaller set of new dims/attrs
  - typically implies dimensional aggregation, not just filtering
  - vocab: projection/mapping

Dimensionality reduction & visualization

- why do people do DR?
  - improve performance of downstream algorithm
  - avoid overfitting
- look for meaning in scatterplots
- includes orthographic projection
- abstract tasks when visualizing DR data
  - synthetic dims created by algorithm
  - naming synthesized dims, mapping synthesized dims to original dims
  - clearing/dimensions, naming clusters, matching clusters and classes

Dimension-oriented tasks

- naming synthesized dims: inspect data represented by lowD points

Cluster-oriented tasks

- verifying, naming, matching to classes

VDA with DR example: nonlinear vs linear

- DR for computer graphics reflectance model
  - goal: simulate how light bounces off materials to make realistic pictures
  - many techniques proposed:
    - many literatures: visualization, machine learning, optimization, psychology, ...
    - principal components analysis (PCA)
    - first try: PCA (linear)
    - result: error falls off sharply after ~45 dimensions
    - many techniques: t-SNE, MDS (multidimensional scaling), charting, tmap, LLE, ...
    - mNE: excellent for clusters
      - but some trickiness remains: http://dl.isi.edu/2014/mnead-paper/
    - MDS: confusing, entire family of techniques, both linear and nonlinear
      - minimize stress or strain metrics
    - early formulations equivalent to PCA

Capturing & using material reflectance

- reflectance measurement: interaction of light with real materials (spheres)
- result: 104 high-res images of material
  - each image 4M pixels
  - goal: image synthesis
  - simply simulate new materials
  - need for more concise model
  - 104 materials * 4M pixels = 400M dims

Understanding synthetic dimensions

- look for meaning in scatterplots
  - synthetic dims created by algorithm but named by human analysts
  - points represent real-world images (spheres)
  - people inspect images corresponding to points to decide if sets could have meaningful name
  - cross-check meaning
    - arrows show simulated images (teaspoons) made from model
    - check if those match dimension semantics

Further reading

  - Chap 13: Reduce Items and Attributes
Ch 14. Embed: Focus+Context

- combine information within single view
- elide
  - selectively filter and aggregate
  - superimpose layer
    - local lens
  - distortion design choices
    - region shape: radial, rectilinear, complex
    - how many regions: one, many
    - region extent: local, global
    - interaction metaphor

Ch 14. Embed: Focus+Context


Ch 15. Analysis Case Studies

- table: draw pixels, sorted by relevance
- group by attribute or partition by attribute into multiple views


**VisDB Analysis**
- Visual Encoding
  - What: Data: Table (clustering with an attribute, query wording example)
  - What Derived: Derived network of nodes and links (roll-up into two chosen attributes)
  - Why: Tasks
  - How: Encode
  - How: Facet
  - How: Paste
  - How: Reduce
  - How: Filter

**Frequency**
- Scale: Nodes by edge frequency

**HCE Analysis**
- Hierarchical Clustering Explorer
  - heatmap, dendrogram
  - multiple views

**PivotGraph Analysis**
- PivotGraph
  - derived rollup network
  - data
  - multi-level network
  - node: word
  - link: words used in same dictionary definition

**Design Study Methodology**
- Reflections from the Trenches and from the Stacks
- Michael Sedlmair
- Tamara Munzner
- Miriah Meyer
- Michael Sedlmair

**Analysis example: Constellation**
- data
  - multi-level network
  - node: word
  - link: words used in same dictionary definition
  - subgraph for each definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale

**HCE**
- rank by feature idiom
  - 1D list
  - 2D matrix

**PivotGraph**
- derived rollup network
- data
- multi-level network
- node: word
- link: words used in same dictionary definition

**Using space: Constellation**
- edge crossings (cannot easily minimize instances, since position constrained by spatial encoding)
- visual encoding

**Constellation Analysis**
- System: Constellation
  - Data: Three-level network of paths, subgraphs (definitions), and nodes (word-senses)
  - Why: Tasks: Discoverability, know and locate types of paths, identity and compare
  - How: Encode: Complete and check links among different spatial positions for plausibility, vertical spatial position for order within path, color links by type
  - How: Manipulate
  - How: Reduce
  - Scale
Methodology for problem-driven work

- definitions
- 9-stage framework
- 32 pitfalls & how to avoid them
- comparison to related methodologies

9-stage framework learning, winnowing, casting, designing, implementing, deploying, reflecting, iterative

Design study methodology: 32 pitfalls

- and how to avoid them

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-1</td>
<td>Premature advance—jumping forward too fast</td>
</tr>
<tr>
<td>PF-2</td>
<td>Premature setup—subjective knowledge of vs. literature</td>
</tr>
<tr>
<td>PF-3</td>
<td>Premature commitment—commitments that are too tight</td>
</tr>
<tr>
<td>PF-4</td>
<td>Prevent too much too fast</td>
</tr>
<tr>
<td>PF-5</td>
<td>Insufficient time available from potential collaborators</td>
</tr>
<tr>
<td>PF-6</td>
<td>No need for visualization; problem can be automated</td>
</tr>
<tr>
<td>PF-7</td>
<td>Research expertise does not match domain problem</td>
</tr>
<tr>
<td>PF-8</td>
<td>No need for research engineering vs. research project</td>
</tr>
<tr>
<td>PF-9</td>
<td>No need for change; existing tools are good enough</td>
</tr>
</tbody>
</table>

Design study methodology: definitions

- 9-stage framework (Precondition, Conceptualization, Core, Analysis)

Considers

- interesting problem
- have data
- have time
- have need
- are you a user?

Roles

- collaborator
- user

Collaborator winnowing

- initial conversation
- further meetings
- prototyping
**Design study methodology: 32 pitfalls**

**PITFALL**

**Premature Design Commitment**

Of course they need the cool technique I built last year!

**Example from the Trenches**

Premature Collaboration!

- Fellow tool builders
- Data promised

**Horse Race vs. Music Debut**

technique-driven problem-driven

Must be first!

Am I ready?

**“writing is research”**

[Wolcott: Writing up qualitative research, 2009]
Cardinality
- Marshall: 68 cities * 40 years * 4 crime types = 10,880
- Wine: 130K * 4 = 650,000

Reflections from the stacks: Wholesale adoption inappropriate
- ethnography
- rapid, goal-directed fieldwork
- grounded theory
- not empty slate vis-a-vis background is key
- action research
- aligned
- intervention as goal
- transferability not reproducibility
- personal involvement is key
- opposition
- translation of participant concepts into visualization language
- researcher need not facilitate design
- art-directed vs vis concern: participants as writers, adversarial to status quo, postmodernity...

Algebraic Process for Visualization Design
- which mathematical structures in data are preserved and reflected in vis
- invariance violation: single dataset, many visualizations
- unambiguity violation: many datasets, same vis
- correspondence violation: doesn’t see change of data in vis
- misreader
- match mathematical structures in data with visual perception
- we can X the data; can we Y the image?
- are important data changes well-matched with obvious visual changes?

Next Steps
- this approach is not the only way to analyze visualizations!
- one specific framework intended to help you think
- other frameworks support different ways of thinking
- following: one interesting example

Visual Design Process In Depth: Dear Data
- data: room occupancy rates
- 1 room
- occupancy measured every 5 min, duration 1 day
- task: characterize space usage patterns
- design
- propose idioms (visual encoding, interaction)
- justify idiom choice

Cardinality
- Marshall: 68 cities * 40 years * 4 crime types = 10,880
- Wine: 130K * 4 = 650,000
- spatial (hierarchical), quantitative, categorical, free-form text

Scenario
- data: room occupancy rates
- 1 room
- occupancy measured every 5 min, duration 1 day
- task: characterize space usage patterns
- design
- propose idioms (visual encoding, interaction)
- justify idiom choice

In-Class Exercise
- http://www.makeovermonday.co.uk/blog/

In-Class Exercise
- http://www.datasketch.es/

Scenario
- what’s the cardinality of the data?
- is a single static chart good enough?
- should you derive any useful additional data?

What-Why-How Analysis
- this approach is not the only way to analyze visualizations!
- one specific framework intended to help you think
- other frameworks support different ways of thinking
- following: one interesting example

Visual Design Process In Depth: Data Sketches
- data: room occupancy rates
- 1 building: 200 rooms across 4 floors
- measured every 5 min, duration 1 day
- task: characterize space usage patterns
- design
- propose & justify idioms

Scenario
- data: room occupancy rates in building
- 1 building: 200 rooms across 4 floors
- measured every 5 min, duration 1 day
- time series + floor plans
- task: characterize space usage patterns
- design
- propose & justify idioms
Consider
• what's the cardinality of the data?
• is a single static chart good enough?
• should you derive any useful additional data?
• how to handle multi-scale space and multi-scale time?
• how to handle multi-scale space and multi-scale time?
• can you normalize the data? should you - always vs on demand?

Scenario
• data: room occupancy rates in building
  – 1 building: 200 rooms across 4 floors
  – measured every 5 min, duration 1 year
  – time series + floor plus + room sizes
  – task: characterize space usage patterns
    – trends, outliers
  – design
    – propose & justify idioms

Consider
• what's the cardinality of the data?
• is a single static chart good enough?
• should you derive any useful additional data?
• what are trade-offs between
  – filtering to see one chart at a time
  – showing side by side with small multiples
    – superimposing on top of each other
• multi-scale structure to exploit: aggregate, zoom, slice/dice, filter?
• can you normalize the data? should you - always vs on demand?
• how to handle multi-scale space and multi-scale time?
• is spatial information germane or extraneous?
• should you normalize the data? should you - always vs on demand?
• how to handle multi-scale space and multi-scale time?
• can you derive any useful additional data?
• what are trade-offs between
  – filtering to see one chart at a time
  – showing side by side with small multiples
    – superimposing on top of each other
• multi-scale structure to exploit: aggregate, zoom, slice/dice, filter?
• can you normalize the data? should you - always vs on demand?
• how to handle multi-scale space and multi-scale time?

Scenario
• data: many metrics across many machines
  – 100 machines, belonging to 20 companies
  – 4 metrics measured every 5 min, duration 1 month
  – CPU, memory, disk I/O, network traffic
  – time series + company name + company location (country)
  – task: forensic analysis to determine possible causes of crashes
  – design
    – propose & justify idioms