Visualization Analysis & Design

I & II

Tamara Munzner
Department of Computer Science
University of British Columbia

Outline
• Session 1 11:15am-2:45pm
  – Analysis: What, Why How
    – Plots and Channels
    – Arrange Tables
    – Arrange Spatial Data
    – Arrange Networks and Trees

• Session 2 2:15pm-3:45pm
  – Map Color and Other Channels
    – Manipulate: Change, Select, Navigate
    – Facet: Juxtapose, Portion, Superimpose
    – Reduce: Filter, Aggregate – Q&A

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

Why?

Why have a human in the loop?
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively. Rather than replace people with computational decision making methods,
• don’t need vis when fully automatic solution exists and is trusted
• many analysis problems ill-specified
  – don’t know exactly what questions to ask in science
• possibilities
  – long-term use for end users (e.g. explanatory analysis of scientific data)
  – presentation of known results
  – stepping stone to better understanding of requirements before developing models
• help developers of automatic solution refine/debug, determine parameters
• help end-users of automatic solutions verify build trust

Why have a human in the loop?

Why a task?

What?

Why use a human vis?

Why use an external representation?
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
• external representation: replace cognition with perception

Why represent all the data?
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
• summaries lose information, details matter
  – confirm expected and find unexpected patterns
  – assess validity of statistical model

Analysis framework: Four levels, three questions
• domain situation
  – who are the target users?
• abstraction
  – translate from specific to domain to vocabulary of vis
    – what is shown? data abstraction
      – domain situation: task
        – why is the user looking at it task abstraction
        – idiom
          – how is it shown?
            – visual encoding idiom
            – interaction idiom
      – algorithms
        – efficient computation

Why is validation difficult?
• solution: use methods from different fields at each level

Why is validation difficult?

Why is validation difficult?

Why validate vis?

Why is validation difficult?

Why is validation difficult?

Why is validation difficult?

Three major datatypes

Attribute types

What?

Actions: Analyze, Query

Why?

Analysis example: Derive one attribute

• don’t just draw what you’re given!
  – decide what the right thing to show is
    – create it with a series of transformations from the original dataset
    – draw that
  • one of the four major strategies for handling complexity

Analysis example: Derive one attribute

• don’t just draw what you’re given!
  – decide what the right thing to show is
    – create it with a series of transformations from the original dataset
    – draw that
  • one of the four major strategies for handling complexity

Analysis example: Derive one attribute

• don’t just draw what you’re given!
  – decide what the right thing to show is
    – create it with a series of transformations from the original dataset
    – draw that
  • one of the four major strategies for handling complexity
Accuracy: Fundamental Theory

25

Channels: Expressiveness types and effectiveness rankings

Magnitude Channels: Ordered Attributes
Identity Channels: Categorical Attributes
Spatial Regions: Geometric primitives
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)

Why: Targets

17

Trends
All Data
Outliers
Attributes
One Many
Distribution
Dependence
Correlation
Similarity
Extremes
Network Data
Spatial Data
Shape
Topology
Paths

How?

Encode Manipulate Facet Reduce

Discriminability: How many usable steps?

Why: Targets

Outline

Visual encoding

Hue (Color)
Position
Length (1D size)
Frequency
Direction
Path

Chap 2: What: Data Abstraction

Analysis: What, Why, How

Marks and Channels
Arrange Tables
Arrange Networks and Trees

Reduce: Filter, Aggregate
Facet: Juxtapose, Partition, Superimpose

Map Color and Other Channels


158.

2014.

attribute levels to show

can only be size coded in 1D

can only be size coded in 1D

cannot be size or shape coded

– and

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–

–
**Idioms: pie chart, polar area chart**

- pie chart
  - area marks with single channel
  - accuracy: angles less accurate than line length
  - length also less accurate than line length
- polar area chart
  - area marks with length channel
  - more direct analogy to bar charts
- data
  - 1 using key attrib. 1 quant value attrib
  - part-to-whole judgements

**Idioms: normalized stacked bar chart**

- task
  - part-to-whole judgements
- normalized stacked bar chart
  - stacked bar chart, normalized to full height
  - single stacked bar equivalent to full pie
  - high information density requires narrow rectangle
- pie chart
  - information density requires large circle

**Idioms: glyph maps**

- rectilinear good for linear vs nonlinear trends
- radial good for cyclic patterns

**Outline**

**Further reading**


**Beware: Population maps trickiness!**

**Idioms: choropleth map**

- use given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic
  - table with 1 quant attribute per region
  - encoding
  - use given accuracy for map mark boundaries
  - sequential segmented colormap (more later)

**Outline**

**Further reading**


**Vector and tensor fields**

- data
  - many strata per cell
- idiom families
  - flow glyph
  - purely local
- geometric flow
  - derived data from tracing particle
  - streamline
  - spatial distribution of seed points
  - feature flow
  - global computation to direct features
  - encoder with use of centroids above

**Outline**


**Further reading**


**Idioms: topographic map**

- data
  - geographic geometry
  - scalar spatial field
  - 1 quant attribute per grid cell
- derived data
  - isoline geometry
  - isocontours computed for specific levels of scalar values

**Idioms: isosurfaces, direct volume rendering**

- data
  - scalar spatial field
  - 1 quant attribute per grid cell
- task
  - shape understanding, spatial relationships
- isosurface
  - derived data used; isocontours computed for specific levels of scalar values
  - direct volume rendering
  - transfer function maps scalar values to color; opacity
  - no derived geometry

**Outline**

**Further reading**

**Idiom: force-directed placement**
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position not meaning directly encoded
  - left to interpret meaning
  - proximity meaningful
- sometimes meaningful
  - sometimes arbitrary artifact of layout algorithm
- tendency to group
  - large edges more readily when short
- tasks
  - explore topic: locate paths, clusters
- scalability
  - node density E4N

**Idiom: adjacency matrix view**
- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quant attr
  - weighted edge between nodes
  - 2 using attr: node list x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - 1K nodes, 1M edges

**Idiom: radial node-link tree**
- data
  - tree
  - 1 quant attr as leaf nodes
- encoding
  - area containment marks for hierarchical structure
  - nonlinear orientation
  - size encodes quant attr
- tasks
  - query attribute as leaf nodes
- scalability
  - 1M leaf nodes

**URLs**
- http://www.treevis.net
Ordered color: Rainbow is poor default
• problems
  – perceptually unordered
  – perceptually nonlinear
• benefits
  – fine-grained structure visible and
  nameable
• alternatives
  – large-scale structure fewer hues
  – small-scale structure multiple hues
  (eg viridis, R/python)

Further reading
– Data 12 Map Color and Other Channel
  ColorBrewer: Brewer: http://www.colorbrewer2.org
  http://www.visinf.tuwien.ac.at/color/

Outline
• Session 1 1/13pm-12:43pm
  – Analysis: What, Why, How
  – Marks and Channels
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• Session 2 2/3pm-4:43pm
  – Map Color and Other Channels
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose
  – Reduce: Filter, Aggregate
  – Q&A

Idiom: Re-encode
System: Tableau
• data: tables with many attributes
• task: compare rankings

Idiom: Animated transitions
System: LineUp
• change view point
  – changes which items are visible within view
  – camera metaphor
  – zoom
    – geometric zoom: faster rotations
    – semantic zoom: adapts object representation based on available pixels
  – panning
  – pan/translate
  – rotate
  – especially in 3D
  – constrained navigation
• often with animated transitions
  – often based on selection set

How to handle complexity: 1 previous strategy + 3 more
• change view over time
• faces across multiple views
• reduce items/attributes within single view
• derive new data to show within view
• change view point
  – changes which items are visible within view
  – camera metaphor
  – zoom
    – geometric zoom: faster rotations
    – semantic zoom: adapts object representation based on available pixels
  – panning
  – pan/translate
  – rotate
  – especially in 3D
  – constrained navigation
• often with animated transitions
  – often based on selection set
Partition into views
• how to divide data between views
– split into regions by attributes
– encodes association between items
– split by state into regions
• why juxtapose views?
  – benefits: easy vs memory
  – lower cognitive load to move eyes between 2 views than remembering previous states with single changing view
  – creates display area, 2 views side by side each have only half the area of one view

Coordinate views: Design choice interaction
• encoding: same/different
  – linked highlighting
• data: all/subset/none
• share navigation

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

Further reading
• Chap 1: Margins and How

Outline
• Session 1: 11:15am–12:45pm
  – Analysis: What, Why, How
  – Marks and Channels
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees
  – Q&A

• Session 2: 2:15pm–3:45pm
  – Map Color and Other Channels
  – Manipulate: Change, Select, Navigate
  – Focus Juxtapose, Partition, Superimpose
  – Reduce: Filter, Aggregate

Idiom: Semantic zooming
• visual encoding change
  – colored box
  – sparkline
  – simple chart
  – full chart: axes and tickmarks

System: LiveRAC
• visual encoding change
  – slice
  – show only items matching specific value for given attribute doing slice
  – axis aligned or arbitrary alignment
  – test
  – show only items on for slice of plane from camera
  – project
  – change mathematics of image creation:
    – rendering
    – perspective
    – many others
  – change colors...

Facet
• Idiom: Small multiples
  – encoding: same
  – data: none shared
    – different attributes for node colors
    – (same network layout)
    – (size)
  – navigation: shared
    – bidirectional linking
  – differences
    – viewport
    – (size)
  – overview-detail

System: EDV
• encoding: same
• data: subset shared
• share navigation
• linked highlighting
• design choice interaction
  – benefits: easy vs memory
  – lower cognitive load to move eyes between 2 views than remembering previous states with single changing view
  – creates display area, 2 views side by side each have only half the area of one view

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

System: Improvise
• investigate power of multiple views
  – pushing limits on view count, interaction complexity
  – how many is ok?
• open research questions
  – reorderable lists
  – easy lookup
  – build on chart linked to other encodings

Partition intoviews
• how to divide data between views
  – split into regions by attributes
  – encodes association between items
  – split by state into regions
  – order of splits has major implications for what patterns are visible
  – no strict dividing line
  – view: big/detailed
    – contiguous region in which visually encoded data is shown on the display
  – view: small/multi-pane:
    – object with internal structure that arises from multiple marks
• why juxtapose views?
  – benefits: easy vs memory
  – lower cognitive load to move eyes between 2 views than remembering previous states with single changing view
  – creates display area, 2 views side by side each have only half the area of one view

Partition: List alignment
• single bar chart with grouped bars
  – split by state into regions
  – compare within a single region
  – compare across regions
  – color by price
  – view: big/detailed
  – contiguous region in which visually encoded data is shown on the display
  – view: big/detailed
  – contiguous region in which visually encoded data is shown on the display
• split by state into regions
  – compare within a single region
  – compare across regions
  – color by price
  – view: big/detailed
  – contiguous region in which visually encoded data is shown on the display
  – view: big/detailed
  – contiguous region in which visually encoded data is shown on the display

System: Hive
• viewing change
  – slice
  – show only items matching specific value for given attribute doing slice
  – axis aligned or arbitrary alignment
  – test
  – show only items on for slice of plane from camera
  – project
  – change mathematics of image creation:
    – rendering
    – perspective
    – many others
  – change colors...

Partition into Side-by-Side Views
• single bar chart with grouped bars
  – split by state into regions
  – compare within a single region
  – compare across regions
  – color by price

System: Cerebral
• encoding: same
• data: none shared
  – different attributes for node colors
  – (same network layout)
• navigation: shared
  – bidirectional linking
  – differences
  – viewport
  – (size)
  – overview-detail

Idiom: Linked highlighting
• see how regions of interest in one view are distributed within another
  – powerful and pervasive interaction idiom
  – encoding: different
  – multiform
  – data: all shared

System: Hive
• viewing change
  – slice
  – show only items matching specific value for given attribute doing slice
  – axis aligned or arbitrary alignment
  – test
  – show only items on for slice of plane from camera
  – project
  – change mathematics of image creation:
    – rendering
    – perspective
    – many others
  – change colors...

Partition: List alignment
• single bar chart with grouped bars
  – split by state into regions
  – compare within a single region
  – compare across regions
  – color by price

System: Improvise
• investigate power of multiple views
  – pushing limits on view count, interaction complexity
  – how many is ok?
• open research questions
  – reorderable lists
  – easy lookup
  – build on chart linked to other encodings

Partition intoviews
• how to divide data between views
  – split into regions by attributes
  – encodes association between items
  – split by state into regions
  – order of splits has major implications for what patterns are visible
  – no strict dividing line
  – view: big/detailed
    – contiguous region in which visually encoded data is shown on the display
  – view: small/multi-pane:
    – object with internal structure that arises from multiple marks

Superimposing limits
• few layers, but many lines
– up to a few dozen
– but not hundreds
• superimpose vs juxtapose: empirical study

Partitioning: Recursive subdivision
• different encoding for second-level regions
– choropleth maps

Dynamic visual layering
• interactive, from selection
• very lightweight: hover
• ex: 1-hop neighbors

System: Cerebral

Superimposing layers
• layer: set of objects spread out over region
– each set is visually distinguishable group
– many layers, how to distinguish?
– encode with different, nonoverlapping channels
– two layers achievable, three with careful design
– small static set, or dynamic from many possible!

Superimpose layers
115
• layer: set of objects spread out over region
– each set is visually distinguishable group
– many layers, how to distinguish?
– encode with different, nonoverlapping channels
– two layers achievable, three with careful design
– small static set, or dynamic from many possible!

Further reading

Idiom: histogram
– static item aggregation
– task: find distribution
– data: table
– derived data
  – new table: keys are bins, values are counts

Idiom: boxplot
– static item aggregation
– task: show skewed distribution
– data: table
– derived data
  – new table keys are bins, values are counts

Idiom: DOSFA
– attribute filtering
– encoding: star glyphs

Idiom: dynamic filtering
– item filtering
– browse through tightly coupled interaction
  – alternative to queries that might return for too many or too few

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

Idiom: scented widgets
– augment widgets for filtering to show information scent
  – taste to show whether value in drilling down further vs looking elsewhere
  – concise, in part of screen normally considered control panel

Dimensionality reduction
– attribute aggregation
– true dimensionality of dataset conjectured to be smaller than dimensionality of measurements

Figure 4 demonstrates these density boxplots applied to 100 numbers drawn from each of four distributions with mean 0 and standard deviation 1: a standard normal, a skew-right distribution (Johnson distribution), a uniform distribution, and a skewed left distribution.

Figure 5 shows a number of alternative visualizations for density estimation in the 12 data sets.


Sneak preview: Not covered today

• Rules of Thumb
  – No unjustified 3D
• Power of the plane, dangers of depth
• Occlusion in 2D
  – Resolution over immersion
  – Overview first, zoom and filter, details on demand
  – Function first, form next

Further reading

  – Chap 1: Reduce Items and Attributes

Outline

• Session 1 11:15am-12:45pm
  – Analysis: What, Why, How
  – Marks and Channels
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees
  – Reduce: Filter, Aggregate
  – Embed: Focus+Context
• Session 2 4:15pm-5:50pm
  – Map Color and Other Channels
  – Manipulate: Change, Select, Navigate
  – Face: Juxtapose, Partition, Superimpose
  – Embed: Focus+Aggregate
• Session 3 3:15am-4:45pm
  – Map Color and Other Channels
  – Manipulate: Change, Select, Navigate
  – Face: Juxtapose, Partition, Superimpose
  – Reduce: Fiber, Aggregate
  – Q&A

More Information

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#halfdaycourse15
• book (including tutorial lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook
  – 20% promo code for book+ebook combo: HVN17
• illustrations: Eamonn Maguire
• papers, videos, software, talks, courses
  http://www.cs.ubc.ca/~tmm/papers
  http://www.cs.ubc.ca/~tmm