Visualization Principles

Tamara Munzner
Department of Computer Science
University of British Columbia

VIZBI 2011: Workshop on Visualizing Biological Data
16 Mar 2011

http://www.cs.ubc.ca/~tmm/talks.html#vizbi11
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively
Defining visualization

Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.

- Human in the loop needs the details

Identical statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>x mean</td>
<td>9.0</td>
</tr>
<tr>
<td>x variance</td>
<td>10.0</td>
</tr>
<tr>
<td>y mean</td>
<td>7.50</td>
</tr>
<tr>
<td>y variance</td>
<td>3.75</td>
</tr>
<tr>
<td>x/y correlation</td>
<td>0.816</td>
</tr>
</tbody>
</table>
Defining visualization

Computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively.

- Human in the loop needs the details
- External representation: perception vs cognition
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

• human in the loop needs the details
• external representation: perception vs cognition
• intended task
Defining visualization

computer-based visualization systems provide visual representations of datasets intended to help people carry out some task more effectively

• human in the loop needs the details
• external representation: perception vs cognition
• intended task
• measureable definitions of effectiveness
Visualization design space

• huge space of design alternatives
  – tradeoffs abound

• many possibilities now known to be ineffective
  • avoid random walk through parameter space
  • avoid some of our past mistakes
  • extensive experimentation has already been done

• guidelines continue to evolve
  – we reflect on lessons learned in design studies
  – iterative refinement usually wise
Principles

• know your visual channel types and ranks
• categorical color constraints
• power of the plane
• danger of depth
• resolution beats immersion
• eyes beat memory
• validate against the right threat
Data types

- Tabular
  - Categorical
  - Ordered
    - Ordinal
  - Quantitative
- Relational
- Spatial
Data types

- tabular
  - categorical
    - fruit: apples, oranges
  - ordered
    - ordinal
  - quantitative

- relational

- spatial
Data types

- Tabular
  - Categorical
    - Fruit: apples, oranges
  - Ordered
    - Ordinal
      - Shirt sizes: small, medium, large
  - Quantitative
- Relational
- Spatial
Data types

- Tabular
  - Categorical
    - Fruit: apples, oranges
  - Ordered
    - Shirt sizes: small, medium, large
- Relational
  - Ordinal
- Spatial
  - Quantitative
    - Lengths: 17 inches, 23 inches
Data types

- tabular
  - categorical
    - fruit: apples, oranges
  - ordered
    - ordinal
      - shirt sizes: small, medium, large
    - quantitative
      - lengths: 17 inches, 23 inches
- relational
  - links between table columns
- spatial
Data types

- **tabular**
  - categorical
    - fruit: apples, oranges
  - ordered
    - ordinal
      - shirt sizes: small, medium, large
    - quantitative
      - lengths: 17 inches, 23 inches

- **relational**
  - links between table columns

- **spatial**
  - intrinsic position, not abstract
Visual encoding

- analyze
  showing abstract data dimensions
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions
Image theory

- marks: geometric primitives
  - points
  - lines
  - areas

- visual channels: control appearance of marks
  - position
    - horizontal
    - vertical
    - both
  - color
  - tilt
  - size
  - shape
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions

1: vertical position

mark: line
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions

1: vertical position

2: vertical position, horizontal position

mark: line       mark: point
Visual encoding

- analyze as combination of marks and channels showing abstract data dimensions

1: vertical position
2: vertical position, horizontal position
3: vertical position, horizontal position, color

mark: line mark: point mark: point
Visual encoding

- Analyze as combination of marks and channels showing abstract data dimensions

1: vertical position
2: vertical position, horizontal position
3: vertical position, horizontal position, color
4: vertical position, horizontal position, color, size

mark: line  mark: point  mark: point  mark: point
Visual channel types and rankings

what/where

How much
Visual channel types and rankings

what/where

planar position
color hue
shape
stipple pattern

How much
Visual channel types and rankings

**What/Where**
- planar position
- color hue
- shape
- stipple pattern

**How much**
- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness (black/white)
- color saturation
- stipple density
Visual channel types and rankings

**Categorical**

*What/where*
- planar position
- color hue
- shape
- stipple pattern

**How much**
- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness (black/white)
- color saturation
- stipple density
Visual channel types and rankings

**Categorical**: what/where
- planar position
- color hue
- shape
- stipple pattern

**Ordered**: Ordinal/Quantitative
- How much
  - position on common scale
  - position on unaligned scale
  - length (1D size)
  - tilt, angle
  - area (2D size)
  - curvature
  - volume (3D size)
  - lightness, black/white
  - color saturation
  - stipple density
Visual channel types and rankings

Categorical
what/where

- planar position
- color hue
- shape
- stipple pattern

Ordered: Ordinal/Quantitative
how much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness (black/white)
- color saturation
- stipple density

Grouping

- containment (2D)
- connection (1D)
- similarity (other channels)
- proximity (position)
Visual channel types and rankings

Categorical
what/where

- planar position
- color hue
- shape
- stipple pattern

Relational, Same Category
Grouping

- Containment (2D)
- Connection (1D)
- Similarity (other channels)
- Proximity (position)

Ordered: Ordinal/Quantitative
How much

- position on common scale
- position on unaligned scale
- length (1D size)
- tilt, angle
- area (2D size)
- curvature
- volume (3D size)
- lightness (black/white)
- color saturation
- stipple density
Power of the plane: only position works for all!

Categorical

- What/where
  - Planar position
  - Color hue
  - Shape
  - Stipple pattern

Ordered: Ordinal/Quantitative

- How much
  - Position on common scale
  - Position on unaligned scale
  - Length (1D size)
  - Tilt, angle
  - Area (2D size)
  - Curvature
  - Volume (3D size)
  - Lightness (black/white)
  - Color saturation
  - Stipple density

Relational, Same Category

- Grouping
  - Containment (2D)
  - Connection (1D)
  - Similarity (other channels)

Proximity (position)
Ranking differs for all other channels

**Categorical**
- What/where
  - planar position
  - color hue
  - shape
  - stipple pattern

**Ordered: Ordinal/Quantitative**
- How much
  - position on common scale
  - position on unaligned scale
  - length (1D size)
  - tilt, angle
  - area (2D size)
  - curvature
  - volume (3D size)
  - lightness black/white
  - color saturation
  - stipple density

**Relational, Same Category**
- Grouping
  - Containment (2D)
  - Connection (1D)
  - Similarity (other channels)
  - Proximity (position)
Channel rankings

• effectiveness principle: encode most important attributes with highest ranked channels [Mackinlay 86]

• where do rankings come from?
  – accuracy, discriminability, separability, popout
Accuracy

Stevens’ Psychophysical Power Law

\[ S = I^n \]

- Electric Current: \( n = 3.5 \)
- Color Saturation: \( n = 1.7 \)
- B/W Lightness: \( n = 1.2 \)
- Length: \( n = 1.0 \)
- Area: \( n = 0.7 \)
- Loudness: \( n = 0.67 \)
- Brightness: \( n = 0.5 \)
Accuracy

- position along common scale

- frame increases accuracy [Cleveland 84]
- Weber's Law: relative judgements
  - filled rectangles differ by 1:9
  - white rectangles differ by 1:2
Discriminability: How many usable steps?

• linewidth: only a few
Discriminability: Categorical color constraints

- noncontiguous small regions of color: only 6-12 bins

*Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. Bioinformatics 2007*
Separability vs. integrality
Separability vs. integrality

- **Position**
- **Hue (color)**

**Fully separable**

2 groups each
Separability vs. integrality

- **Position**
  - Hue (color)
  - Fully separable

- **Size**
  - Hue (color)
  - Some interference
  - Difficult to discriminate small items

- 2 groups each

(2 groups each)
Separability vs. integrality

- Position: fully separable
  - Hue (color)
  - Some interference
  - Difficult to discriminate small items
  - Integral percept: area (planar size)
  - 2 groups each

- Size: some/significant interference
  - Hue (color)
  - Size: width
  - Size: height
  - Integral percept: area
  - 2 groups each

- Size: 3 groups
Separability vs. integrality

- **Position**
  - hue (color)
  - fully separable
  - 2 groups each

- **Size**
  - hue (color)
  - some interference
  - difficult to discriminate small items
  - 2 groups each

- **Size: Width**
  - size: height
  - some/significant interference
  - integral percept: area (planar size)
  - 3 groups

- **Red, Green**
  - major interference
  - integral percept: color/hue
  - 4 groups
Separability vs. integrality

- **Position variety**
  - Hue (color)
  - Fully separable
  - Some interference
  - Difficult to discriminate small items
  - 2 groups each

- **Size variety**
  - Hue (color)
  - Some interference
  - Integral percept: area (planar size)
  - 3 groups

- **Size: Width**
  - Size: height
  - Some/significant interference
  - Integral percept: color/hue
  - 4 groups

- **Red green**
  - Major interference

Note: The diagram illustrates the concept of separability and integrality in visual perception, showing how different attributes (position, size, color) affect the ability to distinguish groups or items.
Separability vs. integrality

- Position and hue (color) are fully separable:
  - 2 groups each

- Size and hue (color) have some interference:
  - Some interference
  - 2 groups each

- Size: width and size: height have some/significant interference:
  - Some/significant interference
  - Integral percept: area (planar size)
  - 3 groups

- Red and green have major interference:
  - Major interference
  - Integral percept: color/hue
  - 4 groups
Popout: Most channels

• parallel processing on most channels
  – sufficiently different item noticed immediately, independent of distractor count

• some channels have no popout: serial search required

Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/
Popout: Most channels

• parallel processing on most channels
  – sufficiently different item noticed immediately, independent of distractor count

• some channels have no popout: serial search required

Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/
Popout limits

• only one channel at a time
  – combination searches are serial
    • most channel pairs
    • all channel triplets, etc

• within channel, speed depends on which channel and how different item is from surroundings
  – ‘sufficiently different’: context dependent

Healey. Perception in Visualization
http://www.csc.ncsu.edu/faculty/healey/PP/
Encoding example: Heatmaps vs. curvemaps

- color traditional, but spatial position outranks it

**heatmap**

**curvemap**

courtesy of M. Styczynski from JavaTreeview jtreeview.sourceforge.net/
Curvemap

- shape perception easier for filled framed line charts than colored boxes

Curvemap

• shape perception easier for filled framed line charts than colored boxes

Curvemap

- shape perception easier for filled framed line charts than colored boxes

Dangers of depth

- rankings for **planar** spatial position, not depth!
- we don’t really live in 3D: we **see** in 2.05D
  - up/down and sideways: image plane
    - acquire more info quickly from eye movements
  - away: depth into scene
    - only acquire more info from head/body motion

- further reading

*Visual Thinking for Design (Chap 5). Colin Ware. 2008*
Dangers of depth: difficulties of 3D

- occlusion
- interaction complexity

Dangers of depth: difficulties of 3D

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

Visualizing the Results of Multimedia Web Search Engines.
Mukherjea, Hirata, and Hara. InfoVis 96
Dangers of depth: difficulties of 3D

• text legibility
  – far worse when tilted from image plane

• further reading

*Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays.*
Grossman et al. CHI 2007

*Visualizing the World-Wide Web with the Navigational View Builder.*
Dangers of depth example

• extruded curves: detailed comparisons impossible

Cluster and Calendar based Visualization of Time Series Data.
van Wijk and van Selow, Proc InfoVis 99.
Transformation to suitable abstraction

• derived data: clusters
• multiple views: calendar, superimposed 2D curves

Cluster and Calendar based Visualization of Time Series Data.
van Wijk and van Selow, Proc InfoVis 99.
Dangers of depth: must justify

• 3D legitimate for true 3D spatial data
• 3D needs very careful justification for abstract data
  – enthusiasm in 1990s, but now skepticism
  – be especially careful with 3D for point clouds or networks

WEBPATH—a three dimensional Web history. Frecon and Smith. InfoVis 1999
Resolution beats immersion

- immersion typically not helpful for abstract data
  - do not need sense of presence or stereoscopic 3D
- resolution much more important
  - pixels are the scarcest resource
  - desktop also better for workflow integration
- virtual reality for abstract data very difficult to justify

Development of an information visualization tool using virtual reality.
Kirner and Martins. Symp Applied Computing 2000
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 two states
  - poor for many states with changes everywhere
    - consider small multiples instead

literal  abstract

animation  small multiples

show time with time  show time with space
Small multiples example: Cerebral

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

Why not animation?

• global comparison difficult
Why not animation?

• further reading

*Animation: can it facilitate? Tversky et al.*
Beyond encoding and interaction

• three more levels of design questions
  – different threats to validity at each level
• validate against the right threat

problem: you misunderstood their needs

abstraction: you’re showing them the wrong thing

encoding: the way you show it doesn’t work

algorithm: your code is too slow

A Nested Model for Visualization Design and Validation. Munzner. IEEE InfoVis 2009.
Characterizing problems of real-world users

• identify a problem amenable to vis
  – provide novel capabilities
  – speed up existing workflow

• validation
  – immediate: interview and observe target users
  – downstream: notice adoption rates
Abstracting into operations on data types

- abstract from domain-specific to generic
- operations
  - sorting, filtering, browsing, comparing, finding trend/outlier,
    characterizing distributions, finding correlation...
- data types
  - tables of numbers, relational networks, spatial
  - transform into useful configuration: derived data
- validation
  - deploy in the field and observe usage
Designing visual encoding, interaction techniques

• visual encoding: drawings they are shown
• interaction: how they manipulate drawings
• validation
  – immediate: careful justification wrt known principles
  – downstream: qualitative or quantitative analysis of results
  – downstream: lab study measuring time/error on given task

• focus of this talk
Creating algorithms to execute techniques

- automatically carry out specification
- validation
  - immediate: complexity analysis
  - downstream: benchmarks for system time, memory
Danger of validation mismatch

- cannot show encoding good with system timings
- cannot show abstraction good with lab study

**Problem validate**: observe target users

**Encoding validate**: justify design wrt alternatives

**Algorithm validate**: measure system time

**Encoding validate**: lab study, qualitative analysis

**Abstraction validate**: observe real usage in field
Principles recap

• know your visual channel types and ranks
• categorical color constraints
• power of the plane
• danger of depth
• resolution beats immersion
• eyes beat memory

• validate against the right threat
More information

• vis intro book chapter
  – principles in more depth
  – also, techniques!

  http://www.cs.ubc.ca/~tmm/papers.html#akpchapter

• papers, videos, software, talks, courses
  http://www.cs.ubc.ca/~tmm

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vizbi11