Visualization Analysis & Design

Tamara Munzner
Department of Computer Science
University of British Columbia
Concurrent Session: Computer Science in College and Universities Conference 2014
October 2014, Vancouver, BC

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

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

• human in the loop needs the details
• doesn’t know exactly what questions to ask in advance
• long-term exploratory analysis
• presentation of known results
• stepping stones towards automation: refining, re-organizing
• external representation: replace cognition with perception
• intended task, measurable definitions of effectiveness

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
• tasks serve as constraint on design (as does data)
• idioms do not serve all tasks equally
• challenge: recent tasks from domain-specific vocabulary to abstract forms
• most possibilities ineffective
• validation is necessary but tricky
• increases chances of finding good solutions if you understand full space of possibilities
• what counts as effective?
• novel enable entirely new kinds of analysis
• faster speed up existing workflows

Why focus on tasks and effectiveness?
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• tasks serve as constraint on design (as does data)
• idioms do not serve all tasks equally
• challenge: recent tasks from domain-specific vocabulary to abstract forms
• most possibilities ineffective
• validation is necessary but tricky
• increases chances of finding good solutions if you understand full space of possibilities
• what counts as effective?
• novel enable entirely new kinds of analysis
• faster speed up existing workflows

Three major datatypes

- Datasets
- What?
- Attributes
- Dataset Types
- Data Types
- Data and Dataset Types
- Tables
- Attributes
- Grids
- Positions
- Items
- Positions
- Grid of positions
- Position

Dataset Types

- 13
- Node (item)
- Fields (Continuous)
- Attributes (columns)
- Value in cell
- Cell
- Grid of positions
- Geometry (Spatial)
- Position

Why analyze?

• imposes structure on huge design space
  - scaffolds to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

Why?

• domain situation
  - who are the target users?
  - abstraction
  - translates from specifics of domain to vocabulary of vis
  - what is shown vs. data abstraction
  - why is the user looking at it vs task abstraction

How?

• idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

What?

• efficient computation

Nested model: Four levels of vis design

Why?

• domain situation
  - who are the target users?
  - abstraction
  - translates from specifics of domain to vocabulary of vis
  - what is shown vs. data abstraction
  - why is the user looking at it vs task abstraction

How?

• idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

What?

• efficient computation

Design Study Methodology

Reflections from the Trenches and from the Socks

Design Studies: Lessons learned after 21 of them

MizBee
- genomics
- Car-X-Ray

in-car networks

Overview
- investigative journalism

Methodology for Problem-Driven Work

• definitions
  - exploratory work

• 9-stage framework
  - 32 pitfalls and how to avoid them

Actions: Analyze, Query

• analyze
  - discover vs present
  - explore vs exploit
  - generate
  - site-specific, mental
  - produce
  - annotate, record, derive
  - query

• how much data matters?
  - one, some, all
  - independent choices

Types: Datasets and data

<table>
<thead>
<tr>
<th>Dataset Types</th>
<th>Data Types</th>
<th>Spatial</th>
<th>Attribute Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datasets</td>
<td>Tables</td>
<td>Networks</td>
<td>Categorical</td>
</tr>
<tr>
<td>Datasets</td>
<td>Spatial</td>
<td>Networks</td>
<td>Ordered</td>
</tr>
<tr>
<td>Datasets</td>
<td>Attribute</td>
<td>Networks</td>
<td>Quantitative</td>
</tr>
</tbody>
</table>

Three major datatypes

- Datasets
- Networks
- Spatial
- Attribute Types

Visualization vs computer graphics
  - geometry is design decision

Why?

• domain situation
  - who are the target users?
  - abstraction
  - translates from specifics of domain to vocabulary of vis
  - what is shown vs. data abstraction
  - why is the user looking at it vs task abstraction

How?

• idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

What?

• efficient computation
### Definitions: Marks and channels
- **Marks**: geometric primitives
  - point
  - line
  - area
- **Channels**: central appearance of marks
  - horizontal position
  - vertical position
  - size (area)
  - color hue
  - lightness
  - saturation
  - luminance
  - position on unaligned scale
  - position on aligned scale
  - depth (3D position)
  - motion (1D position)

### Challenges of Color
- **Expressiveness principle**: match channel and data characteristics
- **Effectiveness principle**: encode most important attributes with highest ranked channels

### Channels: Rankings
<table>
<thead>
<tr>
<th>Categorical Color</th>
<th>Limited Number of Discriminable Bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lightness</td>
<td>Human perception built on fine-grained comparisons</td>
</tr>
<tr>
<td>Hue</td>
<td>Noncongruent small regions of color</td>
</tr>
<tr>
<td>Saturation</td>
<td>Perceptual nonlinearity</td>
</tr>
<tr>
<td>Luminance</td>
<td>Narrow bins than you want</td>
</tr>
</tbody>
</table>

### Categorical vs ordered color
- **Ordered color**: Rainbow is poor default
  - Problems: perceptually unordered, perceptually nonlinear
  - Benefits: fine-grained structure visible and readable

- **Categorical color**: limited number of discriminable bins
  - Problems: narrow bins than you want
  - Benefits: fine-grained structure visible and readable

### How to encode: Arrange space, map channels
- **Arrange**: Express, Separate, Map, Aggregated, and aligned attributes
- **Map**: Color, Hue, Saturation, Lightness, Luminance, Position, Depth, Motion, Shape

### Decomposing color
- **First rule of color**: do not talk about color!
  - Color is confusing if treated as monolithic
- **Decompose into three channels**
  - Ordered can show magnitude
  - Saturation
  - Categorical can show identity
  - Hue
- **Channels have different properties**
  - What they convey directly to perceptual system
  - How much they can convey: how many discriminable bins can we use?
Ordered color: Rainbow is poor default

• problems
  – perceptually unordered
  – perceptually nonlinear
• benefits
  – fine-grained structure visible and nameable
• alternatives
  – large-scale structure fewer hues
  – fine structure multiple hues with monotonically increasing luminance [e.g. viridis Riptide]

Unjustified 3D all too common, in the news and elsewhere

• 3D legitimate for true 3D spatial data
• 3D needs very careful justification for abstract data
– derived data: cluster hierarchy
• juxtapose multiple views: calendar, superimposed 2D curves

Visual encoding: 2D vs 3D

• 2D good, 3D better!
– not so fast…

Perspective distortion loses information

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

3D vs 2D bar charts

• 3D bars never a good idea!

Justified 3D: shape perception

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

Justified 3D: Economic growth curve

• derived new data to show within view
• change view over time
• facet across multiple views
• reduce items/attributes within single view

Magnitude Channels: Ordered Attributes
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)

Life in 3D…

• we don’t really live in 3D: we see in 2.05D
– acquire more info on image plane quickly from eye movements
  – acquire more info for depth slower, from head/body motion

3D vs 2D bar charts

• 3D bars never a good idea!

3D vs 2D bar charts

• 3D bars never a good idea!

3D vs 2D bar charts

• 3D bars never a good idea!

3D vs 2D bar charts

• 3D bars never a good idea!

3D vs 2D bar charts

• 3D bars never a good idea!

3D vs 2D bar charts

• 3D bars never a good idea!