Visualization Analysis & Design
Full-Day Tutorial

Tamara Munzner
Department of Computer Science
University of British Columbia
VIS 2017 Tutorial
September 2017, Phoenix AZ

www.cs.ubc.ca/~tmm/talks.html#vad17fullday
@tamaramunzner
Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

[http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday](http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday)
Defining visualization (vis)

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

Why?...
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.

• human in the loop needs the details & no trusted automatic solution exists
  – doesn't know exactly what questions to ask in advance
  – exploratory data analysis
    • speed up through human-in-the-loop visual data analysis
  – present known results to others
  – stepping stone towards automation
    – before model creation to provide understanding
    – during algorithm creation to refine, debug, set parameters
    – before or during deployment to build trust and monitor
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 depend on vision?

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

• human visual system is high-bandwidth channel to brain
  – overview possible due to background processing
    • subjective experience of seeing everything simultaneously
    • significant processing occurs in parallel and pre-attentively

• sound: lower bandwidth and different semantics
  – overview not supported
    • subjective experience of sequential stream

• touch/haptics: impoverished record/replay capacity
  – only very low-bandwidth communication thus far

• taste, smell: no viable record/replay devices
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

Anscombe’s Quartet

Identical statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x mean</td>
<td>9</td>
</tr>
<tr>
<td>x variance</td>
<td>10</td>
</tr>
<tr>
<td>y mean</td>
<td>7.5</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>

https://www.youtube.com/watch?v=DbJyPELmhJc

Same Stats, Different Graphs
Why focus on tasks and effectiveness?

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

• effectiveness requires match between data/task and representation
  – set of representations is huge
  – many are ineffective mismatch for specific data/task combo
  – increases chance 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

• how to validate effectiveness
  – many methods, must pick appropriate one for your context
What resource limitations are we faced with?

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

• computational limits
  – processing time
  – system memory

• human limits
  – human attention and memory

• display limits
  – pixels are precious resource, the most constrained resource
  – information density: ratio of space used to encode info vs unused whitespace
    • tradeoff between clutter and wasting space, find sweet spot between dense and sparse
Why analyze?
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

SpaceTree

TreeJuxtaposer

---


Further reading

  —Chap 1: What’s Vis, and Why Do It?
Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

[http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday](http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday)
Nested model: Four levels of vis design

- **domain situation**
  - who are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
    - **what** is shown? data abstraction
    - **why** is the user looking at it? task abstraction
- **idiom**
  - **how** is it shown?
    - visual encoding idiom: how to draw
    - interaction idiom: how to manipulate
- **algorithm**
  - efficient computation


Why is validation difficult?

- different ways to get it wrong at each level

**Domain situation**
You misunderstood their needs

**Data/task abstraction**
You’re showing them the wrong thing

**Visual encoding/interaction idiom**
The way you show it doesn’t work

**Algorithm**
Your code is too slow
Why is validation difficult?

- solution: use methods from different fields at each level

**What?**

- Dataset Availability
  - Static
  - Dynamic

**Why?**

**How?**

**Attributes**
- Attribute Types
  - Categorical
  - Ordered
    - Ordinal
    - Quantitative

**Data Types**
- Items
- Attributes
- Links
- Positions
- Grids

**Data and Dataset Types**

<table>
<thead>
<tr>
<th>Tables</th>
<th>Networks &amp; Trees</th>
<th>Fields</th>
<th>Geometry</th>
<th>Clusters, Sets, Lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Items (nodes)</td>
<td>Grids</td>
<td>Items</td>
<td>Items</td>
</tr>
<tr>
<td>Attributes</td>
<td>Links</td>
<td>Positions</td>
<td>Positions</td>
<td>Items</td>
</tr>
</tbody>
</table>

**Dataset Types**
- Tables
- Networks
- Fields (Continuous)
- Multidimensional Table
- Trees
- Geometry (Spatial)

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

---

**Note:**
- Why?
- How?
- What?
Three major datatypes

- **Tables**: Attributes (columns) → Items (rows) → Cell containing value

- **Networks**: Node (item) → Link → Node (item)

- **Spatial**: Fields (Continuous) → Grid of positions → Geometry (Spatial) → Position

**Dataset Types**

- **Tables**
- **Networks**
- **Spatial**

- **Multidimensional Table**

- **Trees**

- **Visualization vs computer graphics**
  - geometry is design decision
Dataset and data types

Data and Dataset Types

<table>
<thead>
<tr>
<th>Tables</th>
<th>Networks &amp; Trees</th>
<th>Fields</th>
<th>Geometry</th>
<th>Clusters, Sets, Lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Items (nodes)</td>
<td>Grids</td>
<td>Items</td>
<td>Items</td>
</tr>
<tr>
<td>Attributes</td>
<td>Links</td>
<td>Positions</td>
<td>Attributes</td>
<td>Positions</td>
</tr>
</tbody>
</table>

Data Types

- Items
- Attributes
- Links
- Positions
- Grids

Dataset Availability

- Static
- Dynamic
Attribute types

Attribute Types
- Categorical
- Ordered
  - Ordinal
  - Quantitative

Ordering Direction
- Sequential
- Diverging
- Cyclic
• \{action, target\} pairs
  – discover distribution
  – compare trends
  – locate outliers
  – browse topology
Actions: Analyze

• consume
  – discover vs present
    • classic split
    • aka explore vs explain
  – enjoy
    • newcomer
    • aka casual, social

• produce
  – annotate, record
  – derive
    • crucial design choice
Derive

• 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

Original Data

Derived Data

\[ \text{trade balance} = \text{exports} - \text{imports} \]
Analysis example: Derive one attribute

• Strahler number
  – centrality metric for trees/networks
  – derived quantitative attribute
  – draw top 5K of 500K for good skeleton

Actions: Search, query

• what does user know? Search
  — target, location

• how much of the data matters?
  — one, some, all

<table>
<thead>
<tr>
<th>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lookup</td>
<td></td>
<td>Browse</td>
</tr>
<tr>
<td>Locate</td>
<td></td>
<td>Explore</td>
</tr>
</tbody>
</table>

• independent choices for each of these three levels
  — analyze, search, query
  — mix and match

Query

Identify

Compare

Summarize
Why: Targets

- **All Data**
  - Trends
  - Outliers
  - Features

- **Attributes**
  - One
    - Distribution
    - Extremes
  - Many
    - Dependency
    - Correlation
    - Similarity

- **Network Data**
  - Topology
    - Paths

- **Spatial Data**
  - Shape
### How?

<table>
<thead>
<tr>
<th><strong>Encode</strong></th>
<th><strong>Manipulate</strong></th>
<th><strong>Facet</strong></th>
<th><strong>Reduce</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arrange</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Express</td>
<td>➔ Change</td>
<td>➔ Juxtapose</td>
<td>➔ Filter</td>
</tr>
<tr>
<td>➔ Separate</td>
<td>➔ Select</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Order</td>
<td>➔ Partition</td>
<td></td>
<td>➔ Aggregate</td>
</tr>
<tr>
<td>➔ Align</td>
<td>➔ Navigate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Use</td>
<td>➔ Superimpose</td>
<td></td>
<td>➔ Embed</td>
</tr>
</tbody>
</table>

- **Map**
  - from *categorical* and *ordered* attributes
  - **Color**
    - ➔ Hue
    - ➔ Saturation
    - ➔ Luminance
  - ➔ Size, Angle, Curvature, ...
  - **Shape**
    - ➔ + • ■ △
  - **Motion**
    - *Direction, Rate, Frequency, ...*

- **What?**
- **Why?**
- **How?**
Further reading

  – Chap 2: What: Data Abstraction
  – Chap 3: Why: Task Abstraction


Outline

- **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels

- **Session 2 10:30am-12:10pm**
  Spatial Layout
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees

- **Session 3 2:00-3:40pm**
  Color & Interaction
  - Map Color
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose

- **Session 4 4:15-5:55pm**
  Guidelines & Methods
  - Reduce: Filter, Aggregate
  - Rules of Thumb
  - Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday @tamaramunzner
Visual encoding

• analyze idiom structure
Definitions: Marks and channels

- **marks**
  - geometric primitives

- **channels**
  - control appearance of marks
  - can redundantly code with multiple channels
Visual encoding

• analyze idiom structure
  – as combination of marks and channels

1: vertical position
mark: line

2: vertical position
horizontal position
mark: point

3: vertical position
horizontal position
color hue
mark: point

4: vertical position
horizontal position
color hue
size (area)
mark: point
Channels

Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Spatial region
Color hue
Motion
Shape
Channels: Matching Types

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape

• expressiveness principle
  – match channel and data characteristics
Channels: Rankings

**Magnitude Channels: Ordered Attributes**

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**

- Spatial region
- Color hue
- Motion
- Shape

• **expressiveness principle**
  – match channel and data characteristics

• **effectiveness principle**
  – encode most important attributes with highest ranked channels
Channels: Expressiveness types and effectiveness rankings

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape

- **expressiveness principle**
  - match channel and data characteristics
- **effectiveness principle**
  - encode most important attributes with highest ranked channels
  - spatial position ranks high for both
Accuracy: Fundamental Theory

Steven’s Psychophysical Power Law: $S = I^N$
Accuracy: Vis experiments

Discriminability: How many usable steps?

• must be sufficient for number of attribute levels to show
  – linewidth: few bins
Separability vs. Integrality

Position
+ Hue (Color)

- Fully separable
  - 2 groups each

Size
+ Hue (Color)

- Some interference
  - 2 groups each

Width
+ Height

- Some/significant interference
  - 3 groups total: integral area

Red
+ Green

- Major interference
  - 4 groups total: integral hue
• find the red dot
  – how long does it take?

• parallel processing on many individual channels
  – 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
• many channels: tilt, size, shape, proximity, shadow direction, ...
• but not all! parallel line pairs do not pop out from tilted pairs
Grouping

- containment
- connection

Marks as Links

- **Containment**
  - ![Diagram of containment](image)

- **Connection**
  - ![Diagram of connection](image)

**Identity Channels: Categorical Attributes**

- **Spatial region**
  - ![Spatial region](image)

- **Color hue**
  - ![Color hues](image)

- **Motion**
  - ![Motion](image)

- **Shape**
  - ![Shapes](image)
Relative vs. absolute judgements

• perceptual system mostly operates with relative judgements, not absolute
  – that’s why accuracy increases with common frame/scale and alignment
  – Weber’s Law: ratio of increment to background is constant
    • filled rectangles differ in length by 1:9, difficult judgement
    • white rectangles differ in length by 1:2, easy judgement

Relative luminance judgements

- perception of luminance is contextual based on contrast with surroundings

http://persci.mit.edu/gallery/checkershadow
Relative color judgements

- color constancy across broad range of illumination conditions

http://www.purveslab.net/seeforyourself/
Further reading

  – Chap 5: Marks and Channels
Outline

- **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels

- **Session 2 10:30am-12:10pm**
  Spatial Layout
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees

- **Session 3 2:00-3:40pm**
  Color & Interaction
  - Map Color
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose

- **Session 4 4:15-5:55pm**
  Guidelines & Methods
  - Reduce: Filter, Aggregate
  - Rules of Thumb
  - Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday
@tamaramunzner
How?

Encode

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

Map
- from **categorical** and **ordered** attributes
  - **Color**
    - Hue
    - Saturation
    - Luminance
  - **Size, Angle, Curvature, ...**
    -悒
- **Shape**
  - + • ■ ▲
- **Motion**
  - *Direction, Rate, Frequency, ...*

Manipulate
- **Change**

Facet
- **Juxtapose**

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

What?

Why?

How?
Encode tables: Arrange space

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

Encode
Arrange tables

Express Values

Separate, Order, Align Regions

Separate

Order

Align

Axis Orientation

Rectilinear

Parallel

Radial

Layout Density

Dense

Space-Filling

1 Key
List

2 Keys
Matrix

3 Keys
Volume

Many Keys
Recursive Subdivision
Keys and values

• key
  – independent attribute
  – used as unique index to look up items
  – simple tables: 1 key
  – multidimensional tables: multiple keys

• value
  – dependent attribute, value of cell

• classify arrangements by key count
  – 0, 1, 2, many...

Express Values

→ 1 Key
  List

→ 2 Keys
  Matrix

→ 3 Keys
  Volume

→ Many Keys
  Recursive Subdivision
Idiom: **scatterplot**

- **express values**
  - quantitative attributes
- **no keys, only values**
  - data
    - 2 quant attrs
  - mark: points
  - channels
    - horiz + vert position
- **tasks**
  - find trends, outliers, distribution, correlation, clusters
- **scalability**
  - hundreds of items

Some keys: Categorical regions

- **regions**: contiguous bounded areas distinct from each other
  - using space to *separate* (proximity)
  - following expressiveness principle for categorical attributes
- use ordered attribute to *order* and *align* regions

- **1 Key**
  - List
- **2 Keys**
  - Matrix
- **3 Keys**
  - Volume
- **Many Keys**
  - Recursive Subdivision
Idiom: **bar chart**

- one key, one value
  - data
    - 1 categ attrib, 1 quant attrib
  - mark: lines
- channels
  - length to express quant value
  - spatial regions: one per mark
    - separated horizontally, aligned vertically
    - ordered by quant attrib
      » by label (alphabetical), by length attrib (data-driven)
- task
  - compare, lookup values
- scalability
  - dozens to hundreds of levels for key attrib
LIMITATION: Hard to know rank. What’s the 4\textsuperscript{th} most? The 7\textsuperscript{th}?

[Slide courtesy of Ben Jones]
Separated, Aligned and Ordered

[Slide courtesy of Ben Jones]
Separated but not Ordered or Aligned

LIMITATION: Hard to make comparisons

[Slide courtesy of Ben Jones]
Idiom: **line chart / dot plot**

- one key, one value
  - data
    - 2 quant attrs
  - mark: points
    - line connection marks between them
  - channels
    - aligned lengths to express quant value
    - separated and ordered by key attrib into horizontal regions
- task
  - find trend
  - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
- scalability
  - hundreds of key levels, hundreds of value levels
Idiom: **stacked bar chart**

- one more key
  - data
    - 2 categ attrib, 1 quant attrib
  - mark: vertical stack of line marks
    - glyph: composite object, internal structure from multiple marks
  - channels
    - length and color hue
    - spatial regions: one per glyph
      - aligned: full glyph, lowest bar component
      - unaligned: other bar components
  - task
    - part-to-whole relationship
  - scalability
    - several to one dozen levels for stacked attrib

Idiom: **streamgraph**

- generalized stacked graph
  - emphasizing horizontal continuity
  - vs vertical items
- data
  - 1 categ key attrib (artist)
  - 1 ordered key attrib (time)
  - 1 quant value attrib (counts)
- derived data
  - geometry: layers, where height encodes counts
  - 1 quant attrib (layer ordering)
- scalability
  - hundreds of time keys
  - dozens to hundreds of artist keys
  - more than stacked bars, since most layers don’t extend across whole chart

Choosing bar vs line charts

• depends on type of key attrib
  – bar charts if categorical
  – line charts if ordered
• do not use line charts for categorical key attribs
  – violates expressiveness principle
  • implication of trend so strong that it overrides semantics!
    – “The more male a person is, the taller he/she is”

Chart axes

• labelled axis is critical
• avoid cropping y-axis
  – include 0 at bottom left
  – or slope misleads

• dual axes controversial
  – acceptable if commensurate
  – beware, very easy to mislead!

http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html
Idiom: connected scatterplots

- scatterplot with line connection marks
  - popular in journalism
  - horiz + vert axes: value attribs
  - line connection marks: temporal order
  - alternative to dual-axis charts
    - horiz: time
    - vert: two value attribs

- empirical study
  - engaging, but correlation unclear

Idiom: **Indexed line charts**

- **data:** 2 quant attires
  - 1 key + 1 value

- **derived data:** new quant value attrib
  - index
  - plot instead of original value

- **task:** show change over time
  - principle: normalized, not absolute

- **scalability**
  - same as standard line chart

[California State Revenues](https://public.tableau.com/profile/ben.jones#!/vizhome/CAStateRevenues/Revenues)
Idiom: **Gantt charts**

- one key, two (related) values
  - data
    - 1 categ attrib, 2 quant attribs
  - mark: line
    - length: duration
- channels
  - horiz position: start/end times
  - horiz length: duration
- task
  - emphasize temporal overlaps, start/end dependencies between items
- scalability
  - dozens of key levels
  - hundreds of value levels


Idiom: heatmap

- two keys, one value
  - data
    - 2 categ attribs (gene, experimental condition)
    - 1 quant attrib (expression levels)
  - marks: area
    - separate and align in 2D matrix
      - indexed by 2 categorical attributes
  - channels
    - color by quant attrib
      - (ordered diverging colormap)
  - task
    - find clusters, outliers
  - scalability
    - 1M items, 100s of categ levels, ~10 quant attrib levels

1Key
List

2 Keys
Matrix

Many Keys
Recursive Subdivision
Axis Orientation

- Rectilinear
- Parallel
- Radial
Idioms: scatterplot matrix, parallel coordinates

• scatterplot matrix (SPLOM)
  – rectilinear axes, point mark
  – all possible pairs of axes
  – scalability
    • one dozen attribs
    • dozens to hundreds of items
• parallel coordinates
  – parallel axes, jagged line representing item
  – rectilinear axes, item as point
    • axis ordering is major challenge
  – scalability
    • dozens of attribs
    • hundreds of items

Task: Correlation

- scatterplot matrix
  - positive correlation
    - diagonal low-to-high
  - negative correlation
    - diagonal high-to-low
  - uncorrelated
- parallel coordinates
  - positive correlation
    - parallel line segments
  - negative correlation
    - all segments cross at halfway point
  - uncorrelated
    - scattered crossings


Idioms: **radial bar chart, star plot**

- radial bar chart
  - radial axes meet at central ring, line mark

- star plot
  - radial axes, meet at central point, line mark

- bar chart
  - rectilinear axes, aligned vertically

- accuracy
  - length unaligned with radial
    - less accurate than aligned with rectilinear
Idioms: **pie chart, polar area chart**

- **pie chart**
  - area marks with angle channel
  - accuracy: angle/area less accurate than line length
    - arclength also less accurate than line length

- **polar area chart**
  - area marks with length channel
  - more direct analog to bar charts

- **data**
  - 1 categ key attrib, 1 quant value attrib

- **task**
  - part-to-whole judgements

Idioms: **normalized stacked bar chart**

- **task**
  - part-to-whole judgements

- **normalized stacked bar chart**
  - stacked bar chart, normalized to full vert height
  - single stacked bar equivalent to full pie
    - high information density: requires narrow rectangle

- **pie chart**
  - information density: requires large circle

http://bl.ocks.org/mbostock/3887235
http://bl.ocks.org/mbostock/3886208
http://bl.ocks.org/mbostock/3886394
Idiom: **glyphmaps**

- rectilinear good for linear vs nonlinear trends

- radial good for cyclic patterns

Orientation limitations

• rectilinear: scalability wrt #axes
  • 2 axes best
  • 3 problematic
    – more in afternoon
  • 4+ impossible

• parallel: unfamiliarity, training time

• radial: perceptual limits
  – angles lower precision than lengths
  – asymmetry between angle and length
  • can be exploited!

Layout Density

Dense

dense software overviews

Further reading

  —Chap 7: Arrange Tables


• A Brief History of Data Visualization. Friendly. 2008.
  http://www.datavis.ca/milestones
Outline

• Session 1 8:30-10:10am
  Visualization Analysis Framework
    – Introduction: Definitions
    – Analysis: What, Why, How
    – Marks and Channels

• Session 2 10:30am-12:10pm
  Spatial Layout
    – Arrange Tables
    – Arrange Spatial Data
    – Arrange Networks and Trees

• Session 3 2:00-3:40pm
  Color & Interaction
    – Map Color
    – Manipulate: Change, Select, Navigate
    – Facet: Juxtapose, Partition, Superimpose

• Session 4 4:15-5:55pm
  Guidelines & Methods
    – Reduce: Filter, Aggregate
    – Rules of Thumb
    – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday
@tamaramunzner
Arrange spatial data

Use Given(11,104),(993,972)

- Geometry
  - Geographic
  - Other Derived

- Spatial Fields
  - Scalar Fields (one value per cell)
    - Isocontours
    - Direct Volume Rendering
  - Vector and Tensor Fields (many values per cell)
    - Flow Glyphs (local)
    - Geometric (sparse seeds)
    - Textures (dense seeds)
    - Features (globally derived)
Idiom: choropleth map

• use given spatial data
  – when central task is understanding spatial relationships

• data
  – geographic geometry
  – table with 1 quant attribute per region

• encoding
  – use given geometry for area mark boundaries
  – sequential segmented colormap [more later]
  – (geographic heat map)

http://bl.ocks.org/mbostock/4060606
Population maps trickiness

• beware!
• absolute vs relative again
  • population density vs per capita
• investigate with Ben Jones Tableau Public demo
  • http://public.tableau.com/profile/ben.jones#!/vizhome/PopVsFin/PopVsFin

Are Maps of Financial Variables just Population Maps?
• yes, unless you look at per capita (relative) numbers

[ https://xkcd.com/1138 ]
Idiom: Bayesian surprise maps

• use models of expectations to highlight surprising values
• confounds (population) and variance (sparsity)

[Surprise! Bayesian Weighting for De-Biasing Thematic Maps. Correll and Heer. Proc InfoVis 2016]
Idiom: **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: isocontours computed for specific levels of scalar values

• direct volume rendering
  – transfer function maps scalar values to color, opacity


Vector and tensor fields

• data
  – many attrs per cell

• idiom families
  – flow glyphs
    • purely local
  – geometric flow
    • derived data from tracing particle trajectories
    • sparse set of seed points
  – texture flow
    • derived data, dense seeds
  – feature flow
    • global computation to detect features
      – encoded with one of methods above


Vector fields

- empirical study tasks
  - finding critical points, identifying their types
  - identifying what type of critical point is at a specific location
  - predicting where a particle starting at a specified point will end up (advection)


Idiom: similarity-clustered streamlines

• data
  – 3D vector field

• derived data (from field)
  – streamlines: trajectory particle will follow

• derived data (per streamline)
  – curvature, torsion, tortuosity
  – signature: 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 Spatial Data


Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday
@tamaramunzner
Arrange networks and trees

- **Node–Link Diagrams**
  - Connection Marks
  - ![Networks and Trees]

- **Adjacency Matrix**
  - Derived Table
  - ![Networks and Trees]

- **Enclosure**
  - Containment Marks
  - ![Networks and Trees]
Idiom: **force-directed placement**

- **visual encoding**
  - link connection marks, node point marks

- **considerations**
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short

- **tasks**
  - explore topology; locate paths, clusters

- **scalability**
  - node/edge density $E < 4N$
Idiom: **sfdp** (multi-level force-directed placement)

- **data**
  - original: network
  - derived: cluster hierarchy atop it

- **considerations**
  - better algorithm for same encoding technique
    - same: fundamental use of space
    - hierarchy used for algorithm speed/quality but not shown explicitly
    - (more on algorithm vs encoding in afternoon)

- **scalability**
  - nodes, edges: 1K-10K
  - hairball problem eventually hits


Idiom: **adjacency matrix view**

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

---

**Figure 7.5:** Comparing matrix and node-link views of a five-node network.

a) Matrix view. b) Node-link view. From [Henry et al. 07], Figure 3b and 3a. (Permission needed.)

Matrix views of networks can achieve very high information density, up to a limit of one thousand nodes and one million edges, just like cluster heatmaps and all other matrix views that uses small area marks.

Technique

<table>
<thead>
<tr>
<th>network matrix view</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Types</td>
</tr>
<tr>
<td>network</td>
</tr>
<tr>
<td>Derived Data</td>
</tr>
<tr>
<td>table: network nodes as keys, link status between two nodes as values</td>
</tr>
<tr>
<td>View Comp.</td>
</tr>
<tr>
<td>space: area marks in 2D matrix alignment</td>
</tr>
<tr>
<td>Scalability</td>
</tr>
<tr>
<td>nodes: 1K</td>
</tr>
<tr>
<td>edges: 1M</td>
</tr>
</tbody>
</table>

---

7.1.3.3 Multiple Keys: Partition and Subdivide

When a dataset has only one key, then it is straightforward to use that key to separate into one region.


Connection vs. adjacency comparison

- **adjacency matrix strengths**
  - 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!

Idiom: **radial node-link tree**

- data
  - tree
- encoding
  - link connection marks
  - point node marks
  - radial axis orientation
    - angular proximity: siblings
    - distance from center: depth in tree
- tasks
  - understanding topology, following paths
- scalability
  - 1K - 10K nodes

Idiom: treemap

• data
  – tree
  – 1 quant attrib at leaf nodes

• encoding
  – area containment marks for hierarchical structure
  – rectilinear orientation
  – size encodes quant attrib

• tasks
  – query attribute at leaf nodes

• scalability
  – 1M leaf nodes

Link marks: Connection and containment

• marks as links (vs. nodes)
  – common case in network drawing
  – 1D case: connection
    • ex: all node-link diagrams
    • emphasizes topology, path tracing
    • networks and trees
  – 2D case: containment
    • ex: all treemap variants
    • emphasizes attribute values at leaves (size coding)
    • only trees

Tree drawing idioms comparison

• data shown
  – link relationships
  – tree depth
  – sibling order

• design choices
  – connection vs containment link marks
  – rectilinear vs radial layout
  – spatial position channels

• considerations
  – redundant? arbitrary?
  – information density?
  • avoid wasting space

Further reading

  –Chap 9: Arrange Networks and Trees


Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday
@tamaramunzner
Idiom design choices: Encode

Encode

Why?

How?

What?

→ Arrange
  → Express
  → Order
  → Use

→ Separate
  → Align

→ Map
  from categorical and ordered attributes

→ Color
  → Hue
  → Saturation
  → Luminance

→ Size, Angle, Curvature, ...

→ Shape

→ Motion
  Direction, Rate, Frequency, ...

What? Why? How?
Categorical vs ordered color

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
    • luminance
    • 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?
Luminance

• need luminance for edge detection
  – fine-grained detail only visible through luminance contrast
  – legible text requires luminance contrast!

• intrinsic perceptual ordering

Spectral sensitivity
Opponent color and color deficiency

- perceptual processing before optic nerve
  - one achromatic luminance channel L
  - edge detection through luminance contrast
  - two chroma channels, R-G and Y-B axis
- “color blind” if one axis has degraded acuity
  - 8% of men are red/green color deficient
  - blue/yellow is rare

Designing for color deficiency: Check with simulator

Normal vision

Deuteranope

Protanope

Tritanope

http://rehue.net

Designing for color deficiency: Avoid encoding by hue alone

- redundantly encode
  - vary luminance
  - change shape

[Deuteranope simulation]

Change the shape

Vary luminance

Color deficiency: Reduces color to 2 dimensions

Normal

Protanope

Deuteranope

Tritanope

Designing for color deficiency: Blue-Orange is safe

Bezold Effect: Outlines matter

- color constancy: simultaneous contrast effect

Color/Lightness constancy: Illumination conditions

Image courtesy of John McCann
Color/Lightness constancy: Illumination conditions

Image courtesy of John McCann
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

– alternatives? this afternoon!

[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]
ColorBrewer

- [http://www.colorbrewer2.org](http://www.colorbrewer2.org)
- saturation and area example: size affects salience!
Ordered color: Rainbow is poor default

• problems
  – perceptually unordered
  – perceptually nonlinear

• benefits
  – fine-grained structure visible and nameable


Ordered color: Rainbow is poor default

• problems
  – perceptually unordered
  – perceptually nonlinear

• benefits
  – fine-grained structure visible and nameable

• alternatives
  – large-scale structure: fewer hues


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 [eg viridis R/python]
Viridis

• colorful, perceptually uniform, colorblind-safe, monotonically increasing luminance

https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
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 [eg viridis R/python]
  – segmented rainbows for binned or categorical


Colormaps

- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate
- Diverging

Categorical
Ordered
Sequential
Bivariate
Diverging

Colormaps

- Categorical
- Ordered
  - Sequential
- Bivariate

- Diverging

http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html

use with care!
Colormaps

- Categorical
- Ordered
  - Sequential
  - Diverging
- Bivariate

- color channel interactions
  - size heavily affects salience
    - small regions need high saturation
    - large need low saturation
  - saturation & luminance: 3-4 bins max
    - also not separable from transparency

Map other channels

• size
  – length accurate, 2D area ok, 3D volume poor

• angle
  – nonlinear accuracy
    • horizontal, vertical, exact diagonal

• shape
  – complex combination of lower-level primitives
  – many bins

• motion
  – highly separable against static
    • binary: great for highlighting
  – use with care to avoid irritation

➡️ Size, Angle, Curvature, ...
  ➡️ Length
  ➡️ Angle
  ➡️ Area
  ➡️ Curvature
  ➡️ Volume

➡️ Shape

➡️ Motion
  ➡️ Motion
  Direction, Rate, Frequency, ...
Angle

Sequential ordered line mark or arrow glyph

Diverging ordered arrow glyph

Cyclic ordered arrow glyph
Further reading

• Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014
  – Chap 10: Map Color and Other Channels

• ColorBrewer, Brewer.
  – http://www.colorbrewer2.org

  – http://www.stonesc.com/Vis06


• https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
Outline

- **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels

- **Session 2 10:30am-12:10pm**
  Spatial Layout
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees

- **Session 3 2:00-3:40pm**
  Color & Interaction
  - Map Color
  - **Manipulate: Change, Select, Navigate**
  - Facet: Juxtapose, Partition, Superimpose

- **Session 4 4:15-5:55pm**
  Guidelines & Methods
  - Reduce: Filter, Aggregate
  - Rules of Thumb
  - Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday

@tamaramunzner
## How?

<table>
<thead>
<tr>
<th>Encode</th>
</tr>
</thead>
<tbody>
<tr>
<td>➔ Arrange ➔ Express ➔ Separate</td>
</tr>
<tr>
<td>➔ Order ➔ Align</td>
</tr>
<tr>
<td>➔ Use</td>
</tr>
</tbody>
</table>

- **Map** from categorical and ordered attributes
  - ➔ Color
    - ➔ Hue ➔ Saturation ➔ Luminance
  - ➔ Size, Angle, Curvature, ...
- **Shape**
  - ➔ + ● ■ ▲
- **Motion**
  - Direction, Rate, Frequency, ...

<table>
<thead>
<tr>
<th>Manipulate</th>
<th>Facet</th>
<th>Reduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>➔ Change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Juxtapose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Filter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Select</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Partition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Aggregate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Navigate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Superimpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>➔ Embed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**What?**

**Why?**

**How?**
How to handle complexity: 1 previous strategy + 3 more

- Derive new data to show within view
- Change view over time
- Facet across multiple views
- Reduce items/attributes within single view

**Manipulate**
- Change

**Facet**
- Juxtapose

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

Change over Time

Navigate

Item Reduction

Attribute Reduction

Zoom

Geometric or Semantic

Slice

Pan/Translate

Constrained

Cut

Project

Select

Change over Time

Navigate

Item Reduction

Attribute Reduction

Zoom

Geometric or Semantic

Slice

Pan/Translate

Constrained

Cut

Project
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: **Realign**

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

**System: LineUp**

Idiom: **Animated transitions**

- smooth interpolation from one state to another
  - alternative to jump cuts, supports item tracking
  - best case for animation
  - staging to reduce cognitive load
- example: animated transitions in statistical data graphics


video: vimeo.com/19278444
Idiom: **Animated transitions - visual encoding change**

- smooth transition from one state to another
  - alternative to jump cuts, supports item tracking
  - best case for animation
  - staging to reduce cognitive load

[Stacked to Grouped Bars](http://bl.ocks.org/mbostock/3943967)
Idiom: Animated transition - tree detail

- animated transition
  - network drilldown/rollup

[Collapsible Tree](https://bl.ocks.org/mbostock/4339083)
Idiom: **Animated transition - bar detail**

- example: hierarchical bar chart
  - add detail during transition to new level of detail

[Hierarchical Bar Chart](https://bl.ocks.org/mbostock/1283663)
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**

Navigate: Reducing attributes

- continuation of camera metaphor
  - slice
    - show only items matching specific value for given attribute: slicing plane
    - axis aligned, or arbitrary alignment
  - cut
    - show only items on far slide of plane from camera
  - project
    - change mathematics of image creation
      - orthographic
      - perspective
      - many others: Mercator, cabinet, ...

Further reading

  – Chap 11: Manipulate View


Outline

• **Session 1 8:30-10:10am**  
  Visualization Analysis Framework  
  – Introduction: Definitions  
  – Analysis: What, Why, How  
  – Marks and Channels

• **Session 2 10:30am-12:10pm**  
  Spatial Layout  
  – Arrange Tables  
  – Arrange Spatial Data  
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**  
  Color & Interaction  
  – Map Color  
  – Manipulate: Change, Select, Navigate  
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**  
  Guidelines & Methods  
  – Reduce: Filter, Aggregate  
  – Rules of Thumb  
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday  
@tamaramunzner
Facet

- Juxtapose

- Partition

- Superimpose
Juxtapose and coordinate views

- Share Encoding: Same/Different
  - Linked Highlighting

- Share Data: All/Subset/None

- Share Navigation
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: **cross filtering**

- item filtering
- coordinated views/controls combined
  - all scented histogram bisliders update when any ranges change

System: **Crossfilter**

[http://square.github.io/crossfilter/]
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

## Coordinate views: Design choice interaction

<table>
<thead>
<tr>
<th>Encoding</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>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
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 question
- reorderable lists
  - easy lookup
  - useful when linked to other encodings

Partition into views

• how to divide data between views
  – split into regions by attributes
  – encodes association between items using spatial proximity
  – 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
  – glyph: small/iconic
    • object with internal structure that arises from multiple marks
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

Partitioning: Recursive subdivision

• size regions by sale counts
  – not uniformly
• result: treemap

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 to distinguish?
    - encode with different, nonoverlapping channels
    - two layers achievable, three with careful design
  - small static set, or dynamic from many possible?
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, multiple for global
  - tasks
    - local: maximum, global: slope, discrimination
  - same screen space for all multiples vs single superimposed

Idiom: **Trellis plots**

- superimpose within same frame
  - color code by year

- partitioning
  - split by site, rows are wheat varieties

- main-effects ordering
  - derive value of median for group, use to order
    - order rows within view by variety median
    - order views themselves by site median
Dynamic visual layering

- interactive based on selection
- one-hop neighbour highlighting demos: click vs hover (lightweight)

Further reading

  - Chap 12: Facet Into Multiple Views
Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday  
@tamaramunzner
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
Idiom: **dynamic filtering**

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

*FilmFinder*

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
Idiom: **scented widgets**

- augmented widgets show information scent
  - cues to show whether value in drilling down further vs looking elsewhere
- concise use of space: histogram on slider


Scented histogram bsliders: detailed

Continuous scatterplot

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

Spatial aggregation

• MAUP: Modifiable Areal Unit Problem
  – gerrymandering (manipulating voting district boundaries) is only one example!
  – zone effects

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

– scale effects

Idiom: **boxplot**

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

![Boxplot Example](image)

*[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

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

\[ \text{data: 9D measured space} \]

\[ \text{DR} \]

\[ \text{Malignant} \]

\[ \text{Benign} \]

\[ \text{derived data: 2D target space} \]
Idiom: Dimensionality reduction for documents

**Task 1**

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD data</td>
<td>2D data</td>
</tr>
</tbody>
</table>

**What?**
- Produce
- Derive

**Why?**
- Discover
- Identify

**Task 2**

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D data</td>
<td>Scatterplot Clusters &amp; points</td>
</tr>
</tbody>
</table>

**What?**
- Produce
- Navigate

**Why?**
- Explore
- Select

**Task 3**

<table>
<thead>
<tr>
<th>In</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scatterplot Clusters &amp; points</td>
<td>Labels for clusters</td>
</tr>
</tbody>
</table>

**What?**
- Produce
- Annotate

**Why?**
- Discover
- Identify
Further reading

  —Chap 13: Reduce Items and Attributes


Outline

• **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  – Introduction: Definitions
  – Analysis: What, Why, How
  – Marks and Channels

• **Session 2 10:30am-12:10pm**
  Spatial Layout
  – Arrange Tables
  – Arrange Spatial Data
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**
  Color & Interaction
  – Map Color
  – Manipulate: Change, Select, Navigate
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**
  Guidelines & Methods
  – Reduce: Filter, Aggregate
  – Rules of Thumb
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday  
@tamaramunzner
Rules of Thumb

• No unjustified 3D
  – Power of the plane
  – Disparity of depth
  – Occlusion hides information
  – Perspective distortion dangers
  – Tilted text isn’t legible

• No unjustified 2D

• Eyes beat memory

• Resolution over immersion

• Overview first, zoom and filter, details on demand

• Responsiveness is required

• Function first, form next
Unjustified 3D all too common, in the news and elsewhere

http://viz.wtf/post/137826497077/eye-popping-3d-triangles

http://viz.wtf/post/139002022202/designer-drugs-ht-ducqn
Depth vs power of the plane

- high-ranked spatial position channels: **planar spatial position**
  - not depth!

**Magnitude Channels: Ordered Attributes**

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)

Steven’s Psychophysical Power Law: \( S = l^N \)
No unjustified 3D: Power of the plane

- high-ranked spatial position channels: **planar** spatial position
  – not depth!

**Magnitude Channels: Ordered Attributes**

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)

Steven’s Psychophysical Power Law: \( S = I^N \)

Perceived Sensation vs. Physical Intensity
No unjustified 3D: Danger of depth

- 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

We can only see the outside shell of the world
Occlusion hides information

- occlusion
- interaction can resolve, but at cost of time and cognitive load

Perspective distortion loses information

- 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]
3D vs 2D bar charts

• 3D bars very difficult to justify!
  – perspective distortion
  – occlusion

• faceting into 2D almost always better choice

[http://perceptualedge.com/files/GraphDesignIQ.html]
Tilted text isn’t legible

• text legibility
  – far worse when tilted from image plane

• further reading


No unjustified 3D example: Time-series data

• extruded curves: detailed comparisons impossible

[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]
No unjustified 3D example: Transform for new data abstraction

- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]
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

• constrained navigation steps through carefully designed viewpoints

No unjustified 3D

- 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

No unjustified 2D

• consider whether network data requires 2D spatial layout
  – especially if reading text is central to task!
  – arranging as network means lower information density and harder label lookup compared to text lists

• benefits outweigh costs when topological structure/context important for task
  – be especially careful for search results, document collections, ontologies
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

![Diagram showing literal vs. abstract, animation vs. small multiples, show time with time vs. show time with space]
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
Resolution beats immersion

• immersion typically not helpful for abstract data
  – do not need sense of presence or stereoscopic 3D
  – desktop also better for workflow integration

• resolution much more important: pixels are the scarcest resource

• virtual reality for abstract data difficult to justify thus far
  • but stay tuned with second wave

Overview first, zoom and filter, details on demand

• influential mantra from Shneiderman


• overview = summary
  — microcosm of full vis design problem

→ Query
  → Identify
  → Compare
  → Summarise
Responsiveness is required

• three major categories
  – 0.1 seconds: perceptual processing
  – 1 second: immediate response
  – 10 seconds: brief tasks

• importance of visual feedback
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

Further reading

  – Chap 6: Rules of Thumb

  – Chap 12: We Have Time Requirements

Outline

- **Session 1 8:30-10:10am**
  Visualization Analysis Framework
  - Introduction: Definitions
  - Analysis: What, Why, How
  - Marks and Channels

- **Session 2 10:30am-12:10pm**
  Spatial Layout
  - Arrange Tables
  - Arrange Spatial Data
  - Arrange Networks and Trees

- **Session 3 2:00-3:40pm**
  Color & Interaction
  - Map Color
  - Manipulate: Change, Select, Navigate
  - Facet: Juxtapose, Partition, Superimpose

- **Session 4 4:15-5:55pm**
  Guidelines & Methods
  - Reduce: Filter, Aggregate
  - Rules of Thumb
  - Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday  @tamaramunzner
Design Study Methodology

Reflections from the Trenches and from the Stacks

http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/
Methodology for problem-driven work

• definitions

• 9-stage framework

• 32 pitfalls & how to avoid them

• comparison to related methodologies
Lessons learned from the trenches: 21 between us

- Cerebral genomics
- MizBee genomics
- Pathline genomics
- MulteeSum genomics
- Vismon fisheries management
- QuestVis sustainability
- WiKeVis in-car networks

- MostVis in-car networks
- Car-X-Ray in-car networks
- ProgSpy2010 in-car networks
- RelEx in-car networks
- Cardiogram in-car networks
- AutobahnVis in-car networks
- VisTra in-car networks

- Constellation linguistics
- LibVis cultural heritage
- Caidants multicast
- SessionViewer web log analysis
- LiveRAC server hosting
- PowerSetViewer data mining
- LastHistory music listening
Design study methodology: definitions

- **Task Clarity**: Fuzzy vs. Crisp
- **Information Location**: Head vs. Computer

**Design Study Methodology**
- Suitable
- Not enough data
- Algorithm automation possible
9 stage framework

- Learn
- Winnow
- Cast
- Discover
- Design
- Implement
- Deploy
- Reflect
- Write

PRECONDITION

CORE

ANALYSIS
9-stage framework

learn
winnow
cast

PRECONDITION

discover design implement deploy reflect write

CORE

ANALYSIS

207
9-stage framework

- PRECONDITION
  - learn
  - winnow
  - cast

- CORE
  - discover
  - design
  - implement
  - deploy

- ANALYSIS
  - reflect
  - write
9-stage framework

- guidelines: confirm, refine, reject, propose
9-stage framework

PRECONDITION

CORE

ANALYSIS

learn → winnow → cast → discover → design → implement → deploy → reflect → write

iterative
Design study methodology: 32 pitfalls

• and how to avoid them

<table>
<thead>
<tr>
<th>PF-1</th>
<th>premature advance: jumping forward over stages</th>
<th>general</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-2</td>
<td>premature start: insufficient knowledge of vis literature</td>
<td>learn</td>
</tr>
<tr>
<td>PF-3</td>
<td>premature commitment: collaboration with wrong people</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-4</td>
<td>no real data available (yet)</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-5</td>
<td>insufficient time available from potential collaborators</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-6</td>
<td>no need for visualization: problem can be automated</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-7</td>
<td>researcher expertise does not match domain problem</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-8</td>
<td>no need for research: engineering vs. research project</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-9</td>
<td>no need for change: existing tools are good enough</td>
<td>winnow</td>
</tr>
</tbody>
</table>
I’m a domain expert! Wanna collaborate?

Of course!!!
considerations

Have data?
Have time?
Have need?

Interesting problem?

...
Are you a user???

... or maybe a fellow tool builder?
Metaphor

Winnowing
Collaborator winnowing
Collaborator winnowing

initial conversation

further meetings
Collaborator winnowing

- initial conversation
- further meetings
- prototyping
- full collaboration
Collaborator winnowing

Talk with many, stay with few!
EXAMPLE FROM THE TRENCHES

Premature Collaboration!

PowerSet Viewer
2 years / 4 researchers

WikeVis
0.5 years / 2 researchers
EXAMPLE FROM THE TRENCHES

Premature Collaboration!

PowerSet Viewer
2 years / 4 researchers

WikeVis
0.5 years / 2 researchers

- Fellow tool builders
- Data promised
Design study methodology: 32 pitfalls

<table>
<thead>
<tr>
<th>PF-10</th>
<th>no real/important/recurring task</th>
<th>winnow</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-11</td>
<td>no rapport with collaborators</td>
<td>winnow</td>
</tr>
<tr>
<td>PF-12</td>
<td>not identifying front line analyst and gatekeeper before start</td>
<td>cast</td>
</tr>
<tr>
<td>PF-13</td>
<td>assuming every project will have the same role distribution</td>
<td>cast</td>
</tr>
<tr>
<td>PF-14</td>
<td>mistaking fellow tool builders for real end users</td>
<td>cast</td>
</tr>
<tr>
<td>PF-15</td>
<td>ignoring practices that currently work well</td>
<td>discover</td>
</tr>
<tr>
<td>PF-16</td>
<td>expecting <em>just talking</em> or <em>fly on wall</em> to work</td>
<td>discover</td>
</tr>
<tr>
<td>PF-17</td>
<td>experts focusing on visualization design vs. domain problem</td>
<td>discover</td>
</tr>
<tr>
<td>PF-18</td>
<td>learning their problems/language: too little / too much</td>
<td>discover</td>
</tr>
<tr>
<td>PF-19</td>
<td>abstraction: too little</td>
<td>design</td>
</tr>
<tr>
<td>PF-20</td>
<td>premature design commitment: consideration space too small</td>
<td>design</td>
</tr>
</tbody>
</table>
Of course they need the cool *technique* I built last year!
METAPHOR

Design Space

+ good
○ okay
- poor
METAPHOR

Design Space

your technique...

+ good
○ okay
- poor
METAPHOR

Design Space

know
METAPHOR
Design Space

know
consider
META PHOR

Design Space

know
consider
propose
METAPHOR
Design Space

know
consider
propose
select
Think broad!
<table>
<thead>
<tr>
<th>PF-21</th>
<th>mistaking technique-driven for problem-driven work</th>
<th>design</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF-22</td>
<td>nonrapid prototyping</td>
<td>implement</td>
</tr>
<tr>
<td>PF-23</td>
<td>usability: too little / too much</td>
<td>implement</td>
</tr>
<tr>
<td>PF-24</td>
<td>premature end: insufficient deploy time built into schedule</td>
<td>deploy</td>
</tr>
<tr>
<td>PF-25</td>
<td>usage study not case study: non-real task/data/user</td>
<td>deploy</td>
</tr>
<tr>
<td>PF-26</td>
<td><em>liking</em> necessary but not sufficient for validation</td>
<td>deploy</td>
</tr>
<tr>
<td>PF-27</td>
<td>failing to improve guidelines: confirm, refine, reject, propose</td>
<td>reflect</td>
</tr>
<tr>
<td>PF-28</td>
<td>insufficient writing time built into schedule</td>
<td>write</td>
</tr>
<tr>
<td>PF-29</td>
<td>no technique contribution ≠ good design study</td>
<td>write</td>
</tr>
<tr>
<td>PF-30</td>
<td>too much domain background in paper</td>
<td>write</td>
</tr>
<tr>
<td>PF-31</td>
<td>story told chronologically vs. focus on final results</td>
<td>write</td>
</tr>
<tr>
<td>PF-32</td>
<td>premature end: win race vs. practice music for debut</td>
<td>write</td>
</tr>
</tbody>
</table>
I can write a design study **paper** in a week!

**“writing is research”**

[Wolcott: Writing up qualitative research, 2009]
Metaphor
Horse Race vs. Music Debut

Must be first!
Am I ready?


http://www.alaineknipes.com/interests/violin_concert.jpg

technique-driven

problem-driven
EXAMPLE FROM THE TRENCHES

Don’t step on your own toes!

First design round published

Subsequent work not stand-alone paper

**AutobahnVis 1.0**
[Sedlmair et al., Smart Graphics, 2009]

**AutobahnVis 2.0**
[Sedlmair et al., Information Visualization 10(3), 2011]
Reflections from the stacks: Wholesale adoption inappropriate

- ethnography
  - rapid, goal-directed fieldwork

- grounded theory
  - not empty slate: 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 lead not facilitate design
    - orthogonal to vis concerns: participants as writers, adversarial to status quo, postmodernity
Outline

• **Session 1 8:30-10:10am**  
  Visualization Analysis Framework  
  – Introduction: Definitions  
  – Analysis: What, Why, How  
  – Marks and Channels

• **Session 2 10:30am-12:10pm**  
  Spatial Layout  
  – Arrange Tables  
  – Arrange Spatial Data  
  – Arrange Networks and Trees

• **Session 3 2:00-3:40pm**  
  Color & Interaction  
  – Map Color  
  – Manipulate: Change, Select, Navigate  
  – Facet: Juxtapose, Partition, Superimpose

• **Session 4 4:15-5:55pm**  
  Guidelines & Methods  
  – Reduce: Filter, Aggregate  
  – Rules of Thumb  
  – Design Study Methodology

http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday  
@tamaramunzner
More Information

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vad17fullday

• book page (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/group/infovis
  http://www.cs.ubc.ca/~tmm

@tamaramunzner