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

Why is validation difficult?
• different ways to get it wrong at each level

Domain situation
You misunderstood their needs
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
Data/task abstraction

Why use an external representation?
• external representation: replace cognition with perception

Why represent all the data?
• summaries lose information, details matter
– confirm expected and find unexpected patterns
– assess validity of statistical model

Analysis framework: Four levels, three questions
– domain situation
– how?
– what?

Actions: Analyze, Query
– Analyze
• analyze
  – discover vs present
  – explore vs explain
• query
  – how much data matters!
  – case, cases, all
– independent choices
  – analyze, query (search)

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

Analysis example: Derive one attribute
• feature number
  – centrality metric for trees/networks
  – derived quantitative attribute
• draw 5K of GSE6 for good sickness

Types: Datasets and data
• Dataset Types
  – Tables
  – Networks
  – Spatial
  – Attribute Types
    – Categorical
    – Ordinal
    – Quantitative

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

Why 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

Why is validation difficult?
• different ways to get it wrong at each level

Domain situation
Observe target users using existing tools
Visual encoding/interaction idiom
The way you show it doesn't work
Algorithm
Your code is too slow

Types: Datasets and data
• Dataset Types
  – Tables
  – Networks
  – Spatial
  – Attribute Types
    – Categorical
    – Ordinal
    – Quantitative

Why is validation difficult?
• different ways to get it wrong at each level

Domain situation
You misunderstood their needs
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
Data/task abstraction
Encoding visually with marks and channels
• analyze idiom structure
– as combination of marks and channels

Channels: Matchings Types
• Magnitude Channels: Ordered Attributes
• Identity Channels: Categorical Attributes
Spatial region
Color hue
Motion
Shape
Direction, Rate, Frequency...
from categorical and ordered attributes

Channels: Rankings
• expressiveness principle
– match channel and data characteristics
• effectiveness principle
– encode most important attributes with highest ranked channels

How to encode: Arrange position and region
• Express Values
• Separate, Order, Align Regions
• Area Orientation
• Rectilinear, Polar, Radial

Arranging tables
• how key, one value
– one value
• two quant attributes
– marks points
• data points
– line connection marks between data points
– channels
• aligned lengths to express quant value
– separated and ordered by key strata into horizontal regions
– task
• find trends

Choosing bar vs line charts
• depends on type of key attributes
– bar charts if categorical
– line charts if ordered
• do not use line charts for categorical key attributes
• violations expressiveness principle
• implication of trend so strong that it overrides semantics!
"The more marks a person is, the taller he is"
Categorical color: Discriminability constraints
• noncontiguous small regions of color: only 6-12 bins

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-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

Ordered color: Rainbow is poor default
• problems
  – perceptually unordered
  – perceptually nonlinear
• benefits
  – fine-grained structure visible and nameable
• alternatives
  – large-scale structure fewer hues

How to handle complexity: 3 more strategies
• change view over time
• facet across multiple views
• reduce items/attributes within single view
• derive new data to show within view

More Information
• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#vad16bryan
• book page (including tutorial lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook
  – 20% promo code for book+ebook combo: HVN17
  – Illustrations: Eamonn Murray
• papers, videos, software, talks, full courses
  http://www.cs.ubc.ca/~tmm/vadbook
  http://www.cs.ubc.ca/~tmm/
• grad vis course Jan ’17: CSPC 547,Tue/Thu 3:30
  – students from outside CS are welcome
  http://www.cs.ubc.ca/~tmm/semevcsun17.html

Viso: A New Visualization System for High-level Visualization