Data Visualization Pitfalls to Avoid

Tamara Munzner
Department of Computer Science
University of British Columbia
CBR Arts Meets Science, UBC Centre for Blood Research
Mar 23 2017, Vancouver BC

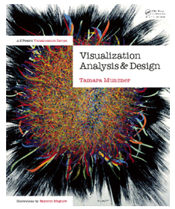
http://www.cs.ubc.ca/~tmm/talks.html#cbr17 @tamaramunzner

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
  - doesn't know exactly what questions to ask in advance
  - longterm exploratory analysis
  - presentation of known results
  - stepping stone towards automation: refining, trustbuilding
- intended task, measurable definitions of effectiveness



more at:
Visualization Analysis and Design, Chapter 1.
Munzner. AK Peters Visualization Series, CRC Press, 2014.

2

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

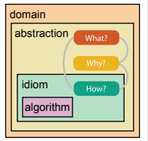
[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008]

3

Nested model: Four levels of vis design

[A Nested Model of Visualization Design and Validation. *Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).*]

- *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



[A Multi-Level Typology of Abstract Visualization Tasks *Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).*]

4

Threats to validity differ 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

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).] 5

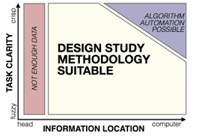
Evaluate success at each level with methods from different fields

anthropology/ ethnography	**Domain situation** Observe target users using existing tools
	Data/task abstraction
design	**Visual encoding/interaction idiom** Justify design with respect to alternatives
computer science	**Algorithm** Measure system time/memory Analyze computational complexity
cognitive psychology	Analyze results qualitatively Measure human time with lab experiment (*lab study*)
anthropology/ ethnography	Observe target users after deployment (*field study*) Measure adoption

problem-driven design studies

technique-driven work

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).] 6



Michael Sedlmair

Miriah Meyer

Design Study Methodology

Reflections from the Trenches and from the Stacks

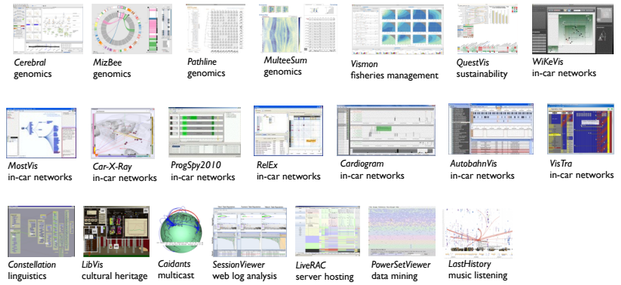
http://www.cs.ubc.ca/labs/imager/tr/2012/dsm/

Tamara Munzner
@tamaramunzner

Design Study Methodology: Reflections from the Trenches and from the Stacks.
Sedlmair, Meyer, Munzner. IEEE Trans. Visualization and Computer Graphics 18(12): 2431-2440, 2012 (Proc. InfoVis 2012).

7

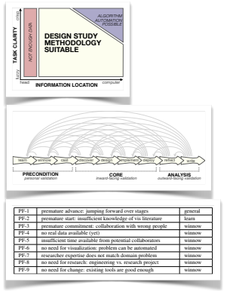
Design Studies: Lessons learned after 21 of them



8

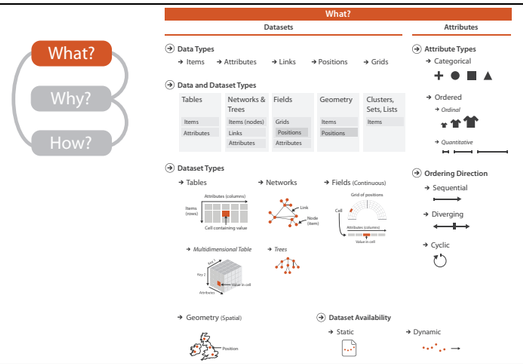
Methodology for Problem-Driven Work

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



PF-1	premature: advance: jumping forward over stages	general
PF-2	premature: start: insufficient knowledge of vis literature	learn
PF-3	premature: commitment: collaborators with wrong people	winnow
PF-4	no real data available (yet)	winnow
PF-5	insufficient time available from potential collaborators	winnow
PF-6	no need for visualization: problem can be automated	winnow
PF-7	researcher expertise does not match domain problem	winnow
PF-8	no need for research: engineering vs. research project	winnow
PF-9	no need for change: existing tools are good enough	winnow

9



10

Three major datatypes

Dataset Types

- Tables
  - Attributes (columns)
  - Items (rows)
  - Cell containing value
  - *Multidimensional Table*
    - Key 1
    - Key 2
    - Value in cell
    - Attributes
- Networks
  - Link
  - Node (item)
  - *Trees*
- Spatial
  - Fields (Continuous)
    - Grid of positions
    - Cell
    - Attributes (columns)
    - Value in cell
  - Geometry (Spatial)
    - Position

- visualization vs computer graphics
  - geometry is design decision

11

Types: Datasets and data

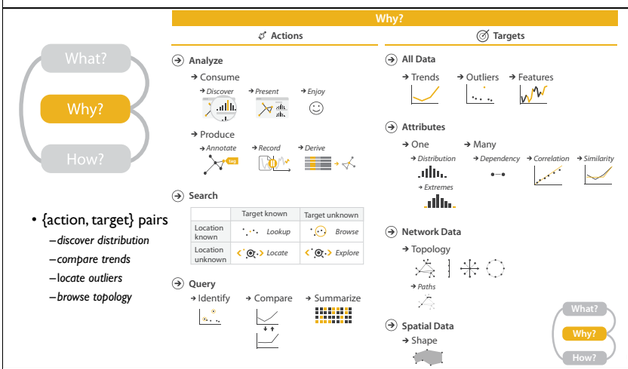
Dataset Types

- Tables
  - Attributes (columns)
  - Items (rows)
  - Cell containing value
- Networks
  - Link
  - Node (item)
- Spatial
  - Fields (Continuous)
    - Grid of positions
    - Cell
    - Attributes (columns)
    - Value in cell
  - Geometry (Spatial)
    - Position

Attribute Types

- Categorical
- Ordered
  - *Ordinal*
  - *Quantitative*

12



- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

13

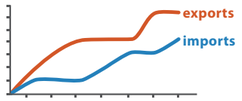
Actions: Analyze, Query

- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social
  - produce
    - annotate, record, derive
  - query
    - how much data matters?
      - one, some, all
- independent choices

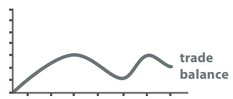
14

Derive: Crucial Design Choice

- 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



trade balance = exports − imports

Derived Data

15

Targets

All Data

- Trends
- Outliers
- Features

Attributes

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

Network Data

- Topology
- *Paths*

Spatial Data

- Shape

16

How?

Encode

- Arrange
  - Express
  - Separate
  - Order
  - Align
  - Use
- Map
  from categorical and ordered attributes
  - Color
    - Hue
    - Saturation
    - Luminance
  - Size, Angle, Curvature, ...
  - Shape
  - Motion
    Direction, Rate, Frequency, ...

Manipulate

- Change
- Select
- Navigate

Facet

- Juxtapose
- Partition
- Superimpose

Reduce

- Filter
- Aggregate
- Embed

What?
Why?
How?

How to encode: Arrange space, map channels

Encode

- Arrange
  - Express
  - Separate
  - Order
  - Align
  - Use
- Map
  from categorical and ordered attributes
  - Color
    - Hue
    - Saturation
    - Luminance
  - Size, Angle, Curvature, ...
  - Shape
  - Motion
    Direction, Rate, Frequency, ...

Definitions: Marks and channels

- marks
  - geometric primitives
- channels
  - control appearance of marks

Points
Lines
Areas

Position
- Horizontal
- Vertical
- Both

Color
Shape
Tilt
Size
- Length
- Area
- Volume

Encoding visually with marks and channels

- 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)

Same

- 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)

Same

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)

Same

Best
Effectiveness
Least

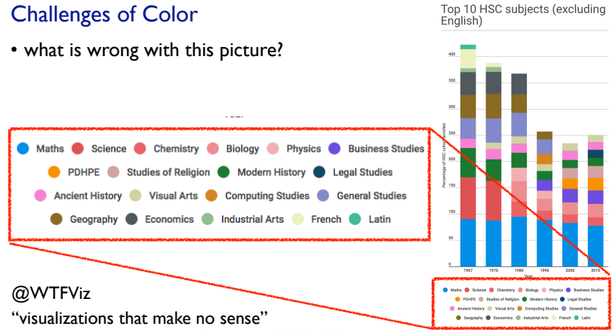
Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

- expressiveness principle
  - match channel and data characteristics
- effectiveness principle
  - encode most important attributes with highest ranked channels

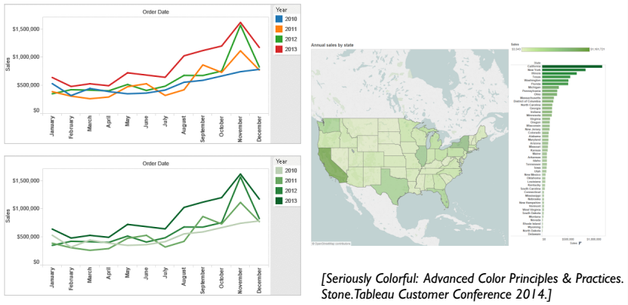
Challenges of Color

- what is wrong with this picture?



@WTFViz
"visualizations that make no sense"
http://viz.wtf/post/150780948819/maths-enrolments-drop-to-lowest-rate-in-50-years

Categorical vs ordered color



[Seriously Colorful: Advanced Color Principles & Practices. Stone. Tableau Customer Conference 2014.]

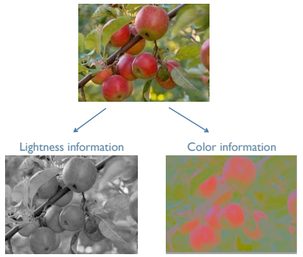
Decomposing color

- 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
Saturation
Hue

Luminance

- need luminance for edge detection
  - fine-grained detail only visible through luminance contrast
  - legible text requires luminance contrast!
- intrinsic perceptual ordering

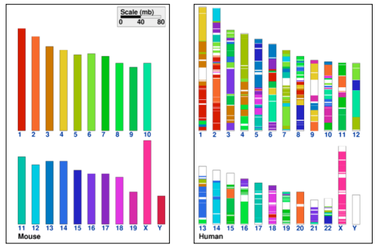


[Seriously Colorful: Advanced Color Principles & Practices. Stone. Tableau Customer Conference 2014.]

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

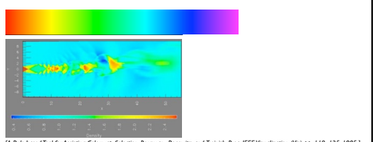
– so what can we do instead?



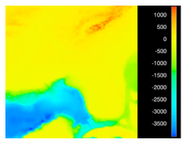
[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable



[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

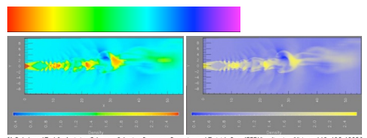


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

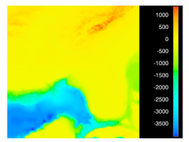
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Ordered color: Rainbow is poor default

- problems
  - perceptually unordered
  - perceptually nonlinear
- benefits
  - fine-grained structure visible and nameable
- alternatives
  - large-scale structure: fewer hues



[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]

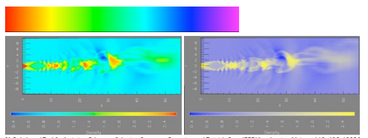


[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

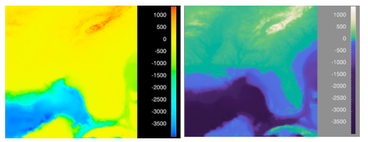
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

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]



[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



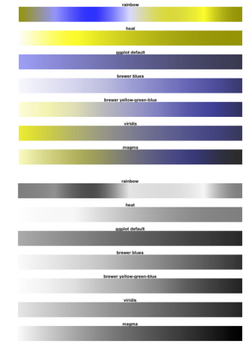
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Viridis

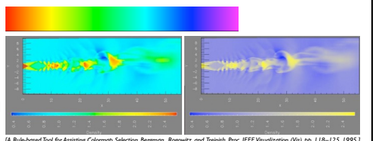
- colorful, perceptually uniform, colorblind-safe, monotonically increasing luminance

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

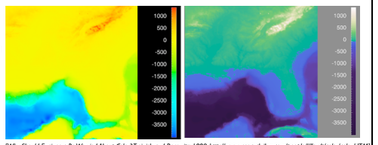


Ordered color: Rainbow is poor default

- 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



[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

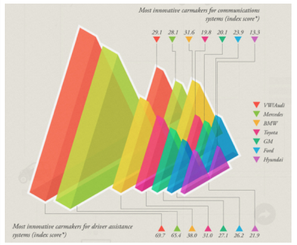
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Visual encoding: 2D vs 3D

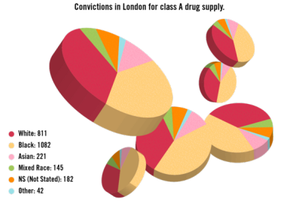
- 2D good, 3D better?
  - –not so fast…

http://amberleyromo.com/images/Bookcover/Animal-Farm.png

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

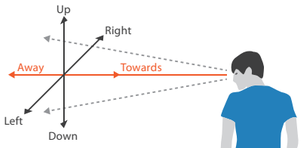
- high-ranked spatial position channels: **planar** spatial position
  - –not depth!

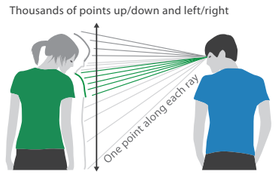
Magnitude Channels: Ordered Attributes

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

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

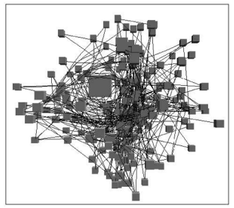




[adapted from Visual Thinking for Design. Ware. Morgan Kaufmann 2010.]

Occlusion hides information

- occlusion
- interaction complexity



[Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.]

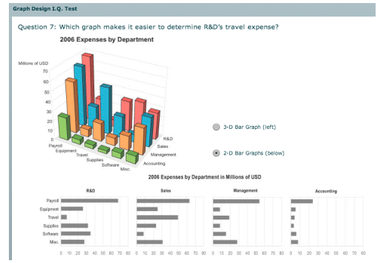
Perspective distortion loses information

- 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

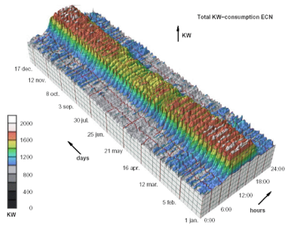
- 3D bars never a good idea!



[http://perceptualedge.com/files/GraphDesignIQ.html]

No unjustified 3D example: Time-series data

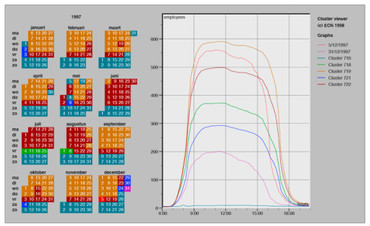
- extruded curves: detailed comparisons impossible



[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

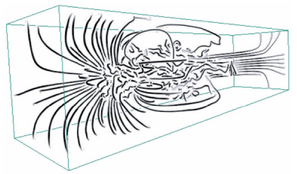
- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves



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

Justified 3D: shape perception

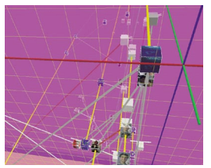
- benefits outweigh costs when task is shape perception for 3D spatial data
  - –interactive navigation supports synthesis across many viewpoints



[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630–640.]

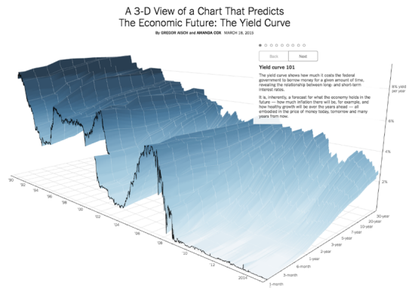
No unjustified 3D

- 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



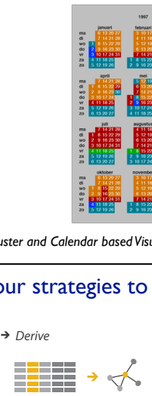
[WEBPATH-a three dimensional Web history. Frecon and Smith. Proc. InfoVis 1999]

Justified 3D: Economic growth curve



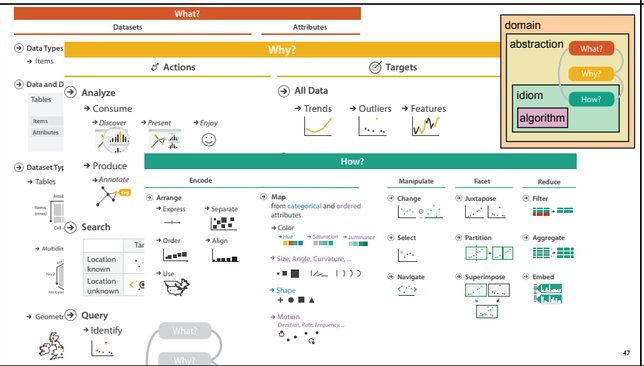
http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

Four strategies to handle complexity



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

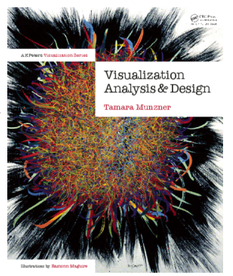
more at:
Visualization Analysis and Design.
Munzner. AK Peters Visualization Series, CRC Press, 2014.



More Information

@tamaramunzner

- this talk
  www.cs.ubc.ca/~tmm/talks.html#cbr17
- book
  http://www.cs.ubc.ca/~tmm/vadbook
  – 20% off promo code, book+ebook combo: HVN17
  – http://www.crcpress.com/product/isbn/9781466508910
- papers, videos, software, talks, courses
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm



Visualization Analysis and Design.
Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.