Introduction to Visualization

part 1

Noeska Smit, Jan Byška *et al.*,
UiB Dept. of Informatics,
2017-08-28

Introduction - The VisGroup at the UiB

Researching and teaching new solutions for the efficient and effective visualization of large and complex datasets

from

- **measurements** (e.g., from medical imaging modalities or from seismic/sonar sensors),
- **computational simulation** (e.g., based on computational fluid dynamics), or from
- **analytic modeling** (e.g., in the form of difference or differential equations)

for the purpose of

- data exploration, analysis, and presentation.

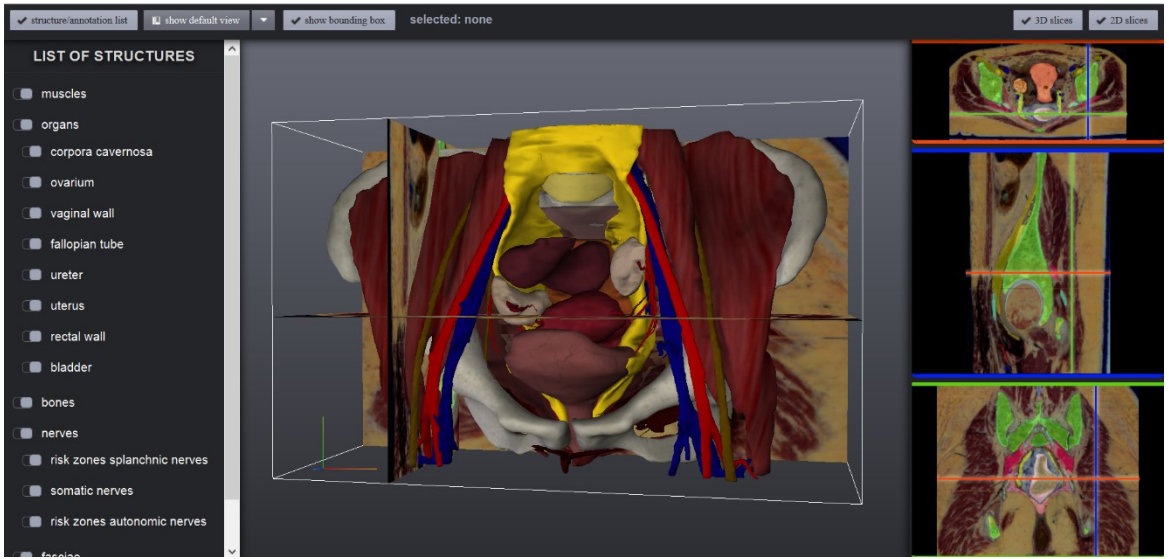
Introduction – Noeska Smit (1)

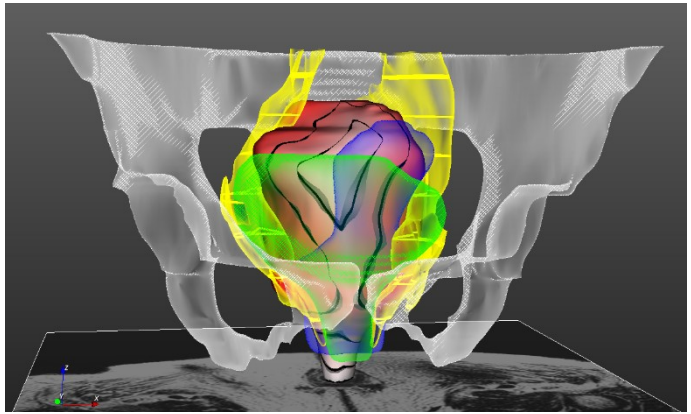
- **Licensed Radiographer**
- **Studied Computer Science in Delft, the Netherlands**
  - Specialized in Computer Graphics & Visualization
- **PhD in Medical Visualization**
- **Currently Associate Professor in the VisGroup**

Noeska Smit Noeska.Smit@UiB.no	**Visualization Group** UiB Dept. of Informatics www.ii.UiB.no/vis

Introduction – Noeska Smit (2)

- **Model-based visualization of human anatomy for medical education and surgical planning**





http://anatomy.tudelft.nl

Introduction Jan Byška (1)

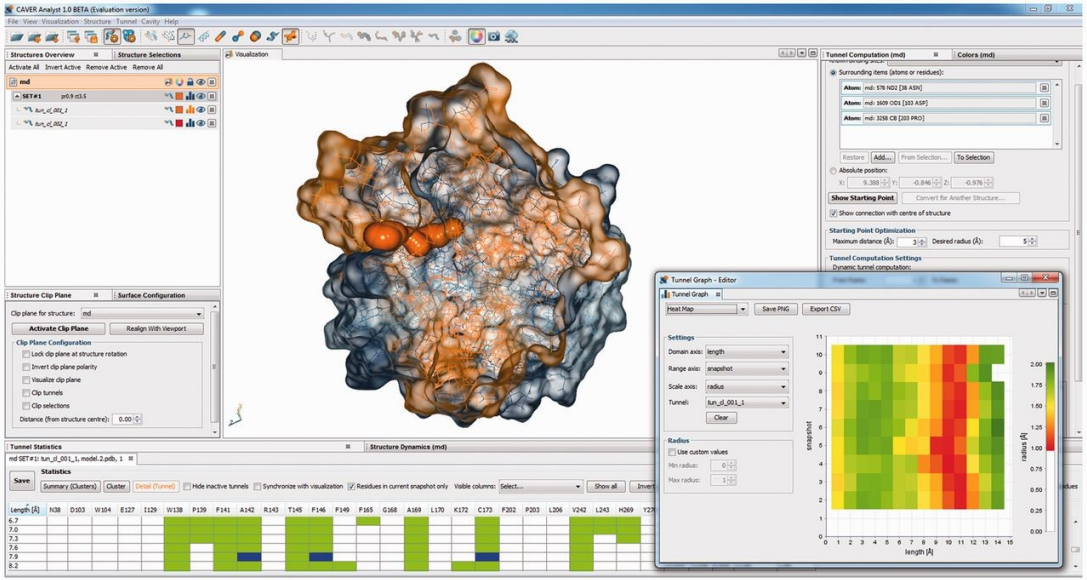
- **Studied Computer Science in Brno, Czech Republic**
  - Specialized in Computer Graphics & Visualization
- **PhD in Molecular Visualization**
- **Currently Postdoc in the VisGroup**

Jan Byška
Jan.Byska@UiB.no
(+47) 96824828

Visualization Group
UiB Dept. of Informatics
www.ii.UiB.no/vis

Introduction Jan Byška (2)

CAVER Analyst - Software tool for analysis and visualization of tunnels and channels in protein structures



http://www.caver.cz/

What is Visualization?

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

The purpose of computing is insight, not numbers

[R. Hamming, 1962]

The purpose of visualization is insight, not pictures

[B. Shneiderman, 1999]

What is visualization?

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

Why?

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

Why have a human in the loop?

No need for vis if trustworthy automatic solution exists

Many analysis problems ill-specified:

- Not sure what questions to ask in advance

Possibilities:

- Long-term use for end users
- Presentations of known results
- Stepping stone to better understand requirements before developing models
- Help developers of automatic solutions refine/debug/determine parameters
- Help end users of automatic solutions verify and build trust

Why use external representation?

Visual representations 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.]

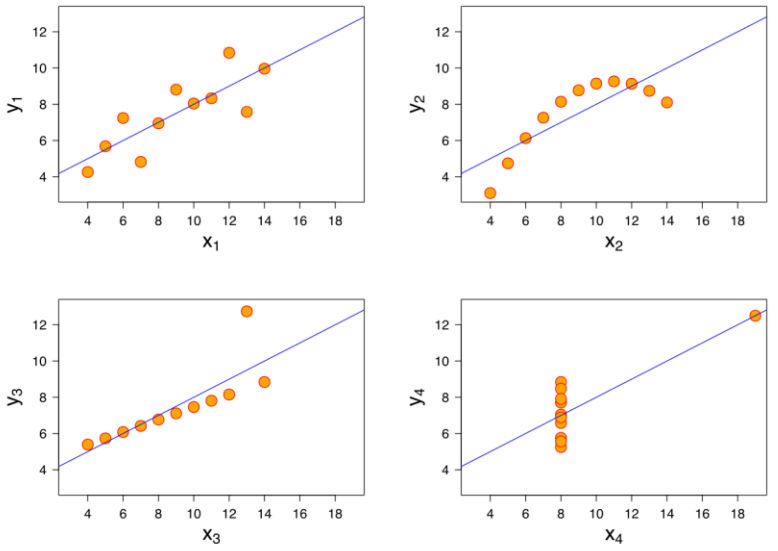
Why represent all the data?

Summaries lose information, details matter

- Confirm expected and find unexpected patterns

Anscombe's quartet, 4 datasets:

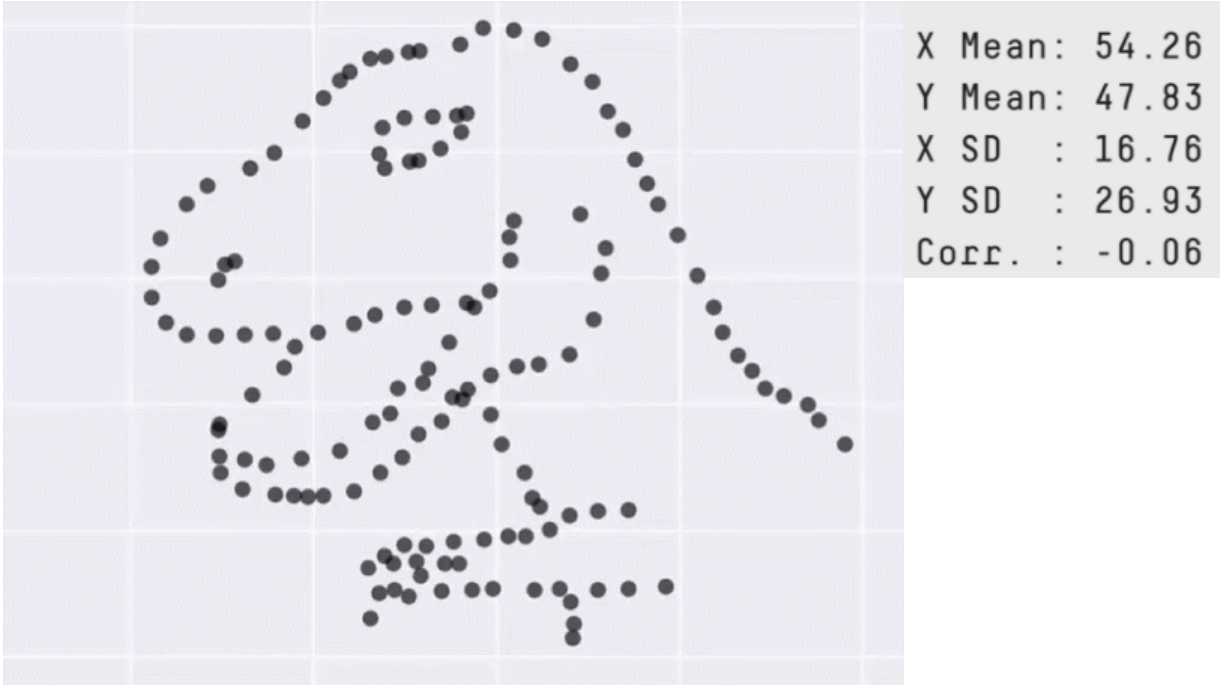
Identical statistics	
x mean	9
x variance	10
y mean	8
y variance	4
x/y correlation	1



https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Same Stats, Different Graphs

The Datasaurus Dozen:



https://www.autodeskresearch.com/publications/samestats

Analysis Framework: Four levels, three questions

Domain situation

- who are the target users?

Abstraction

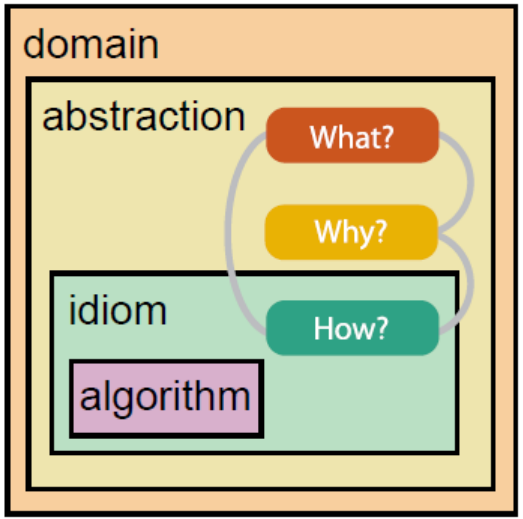
- translate from domain specifics to vis vocabulary
- **what** is shown? data abstraction
- **why** is the user looking at it? task abstraction

Idiom

- **how** is it shown? visual encoding and interaction

Algorithm

- efficient computation

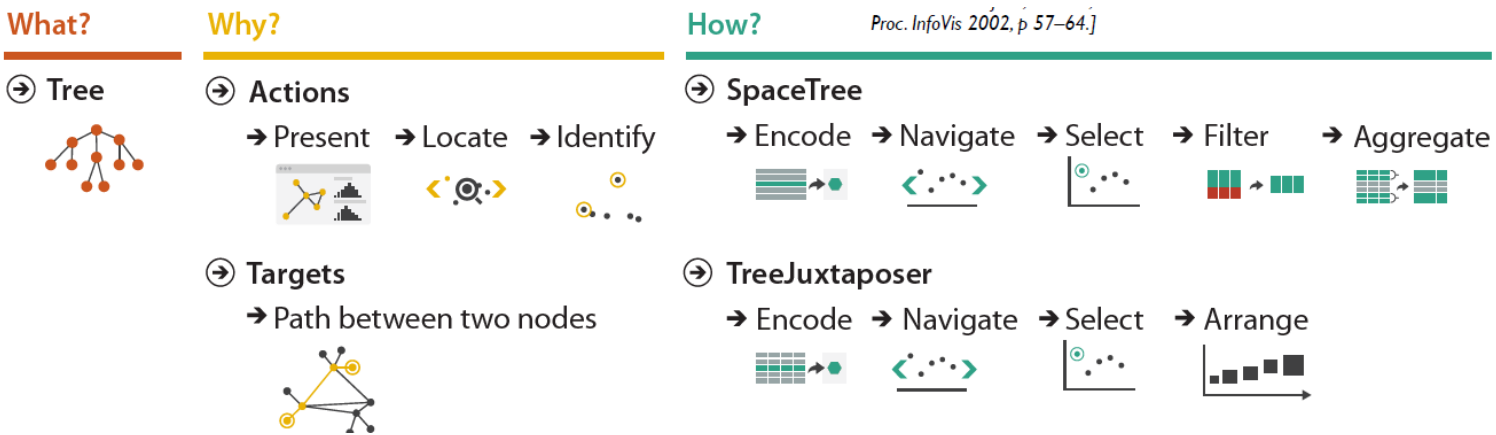


[A Nested Model of Visualization Design and Validation. *Munzner. IEEE TVCG 2009]*
[A Multi-Level Typology of Abstract Visualization Tasks. *Brehmer and Munzner. IEEE TVCG 2013]*

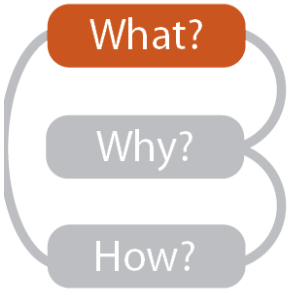
Why analyze?

Structuring visualization design space:

- think systematically about choices
- analyze existing as stepping stone to new



What?



What?

Datasets

Data Types

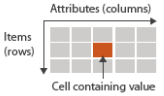
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

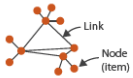
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

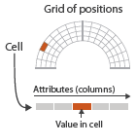
- Tables



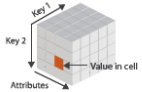
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

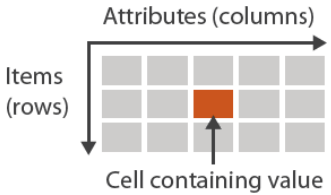
- Diverging

- Cyclic

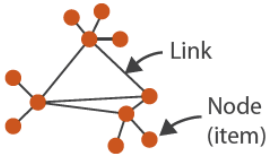
Types: Datasets and Data

Dataset Types

Tables

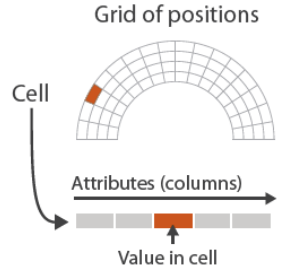


Networks



Spatial

Fields (Continuous)



Geometry (Spatial)

Attribute Types

Categorical

Ordered

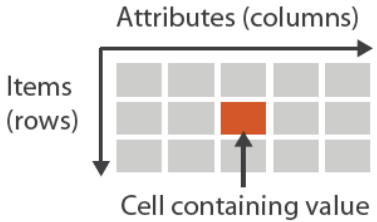
Ordinal

Quantitative

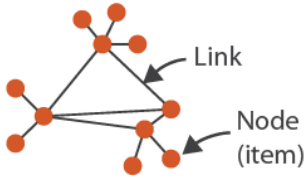
Three major datatypes

Dataset Types

Tables

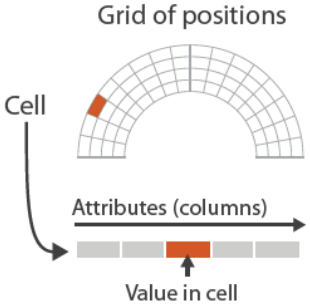


Networks



Spatial

Fields (Continuous)



Geometry (Spatial)

Attribute types and ordering

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

Why?

{action, target} pairs:

- discover distribution
- compare trends
- locate outliers
- browse topology

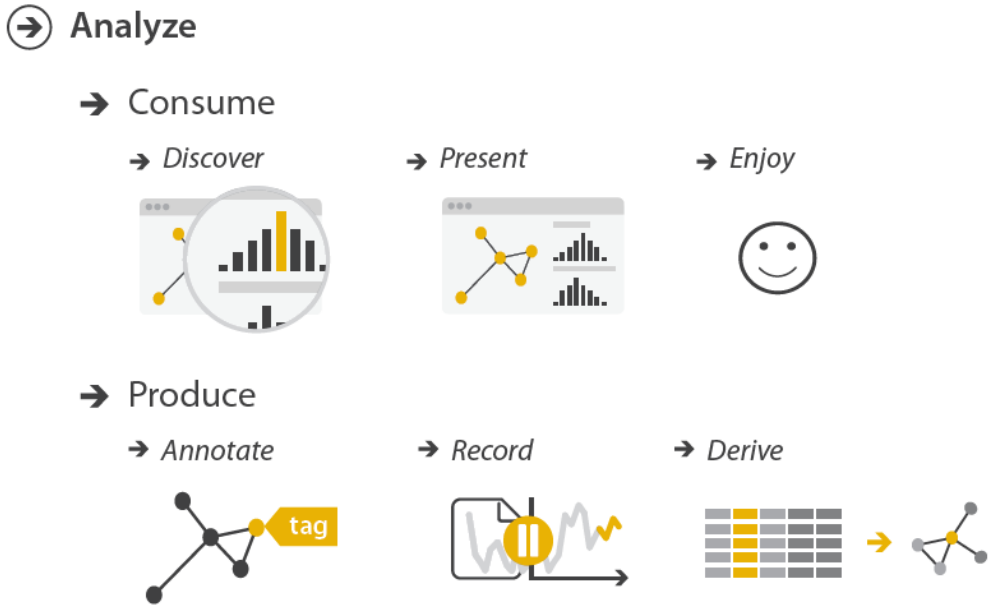
Actions I: Analyze

consume

- discover vs present (classic split, explore vs explain)
- enjoy (newcomer, casual)

produce

- annotate, record
- derive

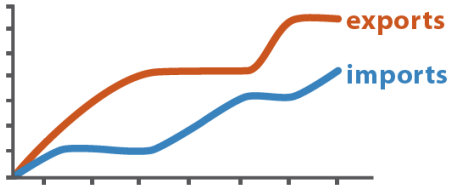


Derive

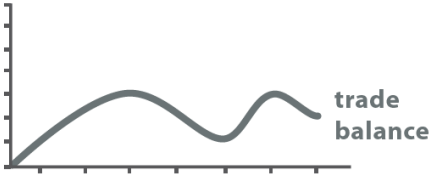
Not just drawing what is given, but:

- deciding what the right thing to show is
- creating it with a series of transformations
- drawing that

One of the major strategies for handling complexity



Original Data



trade balance = exports – imports

Derived Data

Actions II: Search

What does the user know?

– target, location

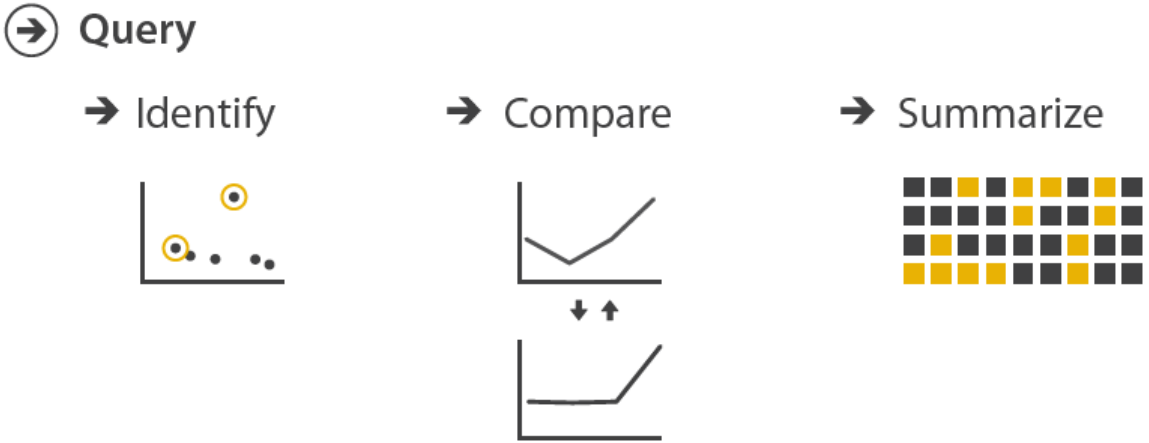
Search

	Target known	Target unknown
Location known	 *Lookup*	 *Browse*
Location unknown	 *Locate*	 *Explore*

Actions III: Query

How much of the data matters?

– one, some, all



analyze, search, query

– independent choices for each

Targets

All Data

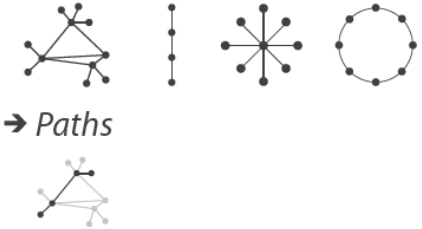
- Trends
- Outliers
- Features

Attributes

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

Network Data

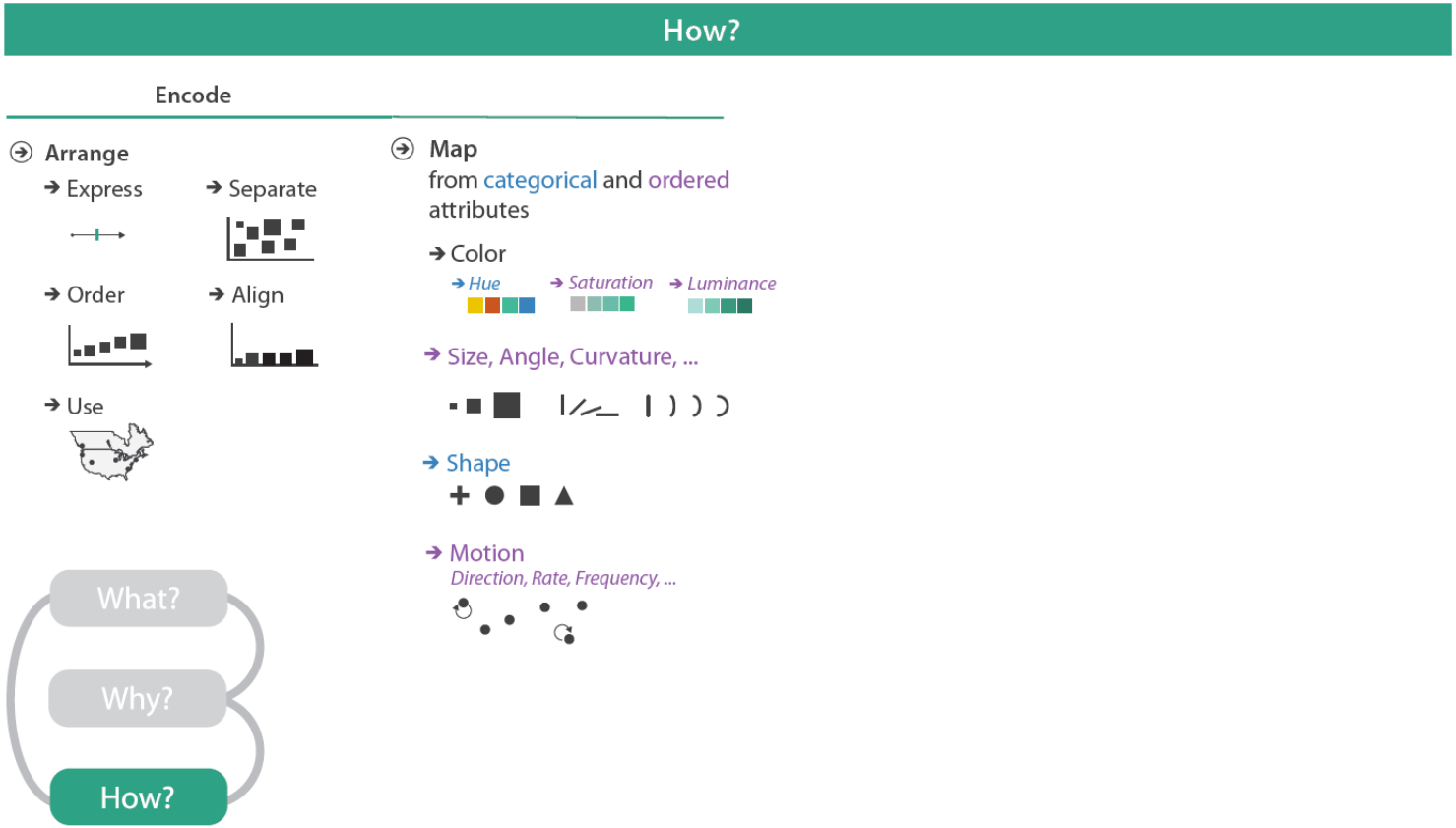
- Topology
- *Paths*



Spatial Data

- Shape

How?


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, ...
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

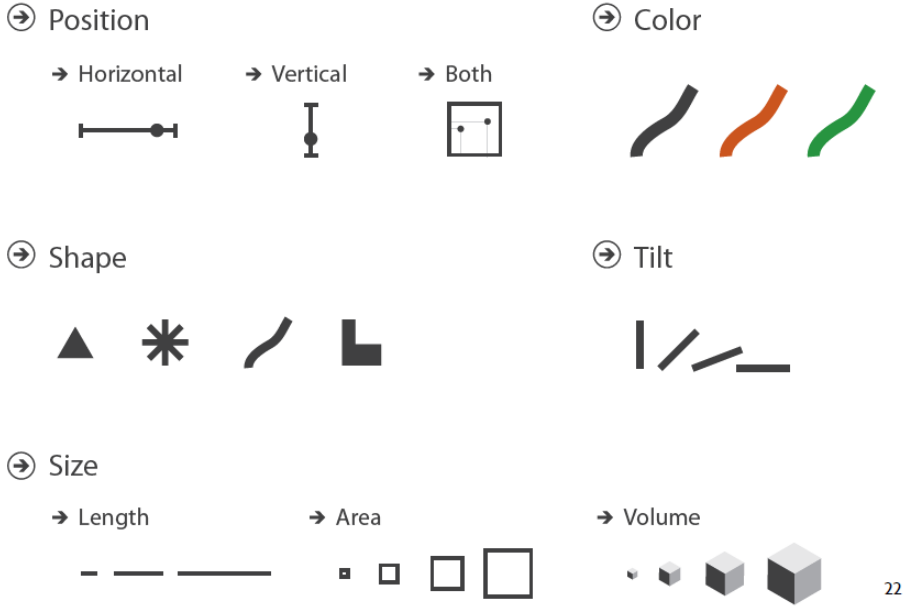
- Motion

  Direction, Rate, Frequency, ...

Definitions: Marks and Channels

- marks
  - geometric primitives

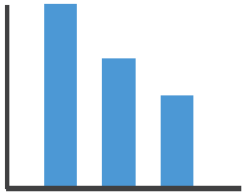
- channels
  - control appearance of marks



Encoding visually with marks and channels

Analyze idiom structure:

- Combination of marks and channels



1:
vertical position

mark: line

2:
vertical position
horizontal position

mark: point

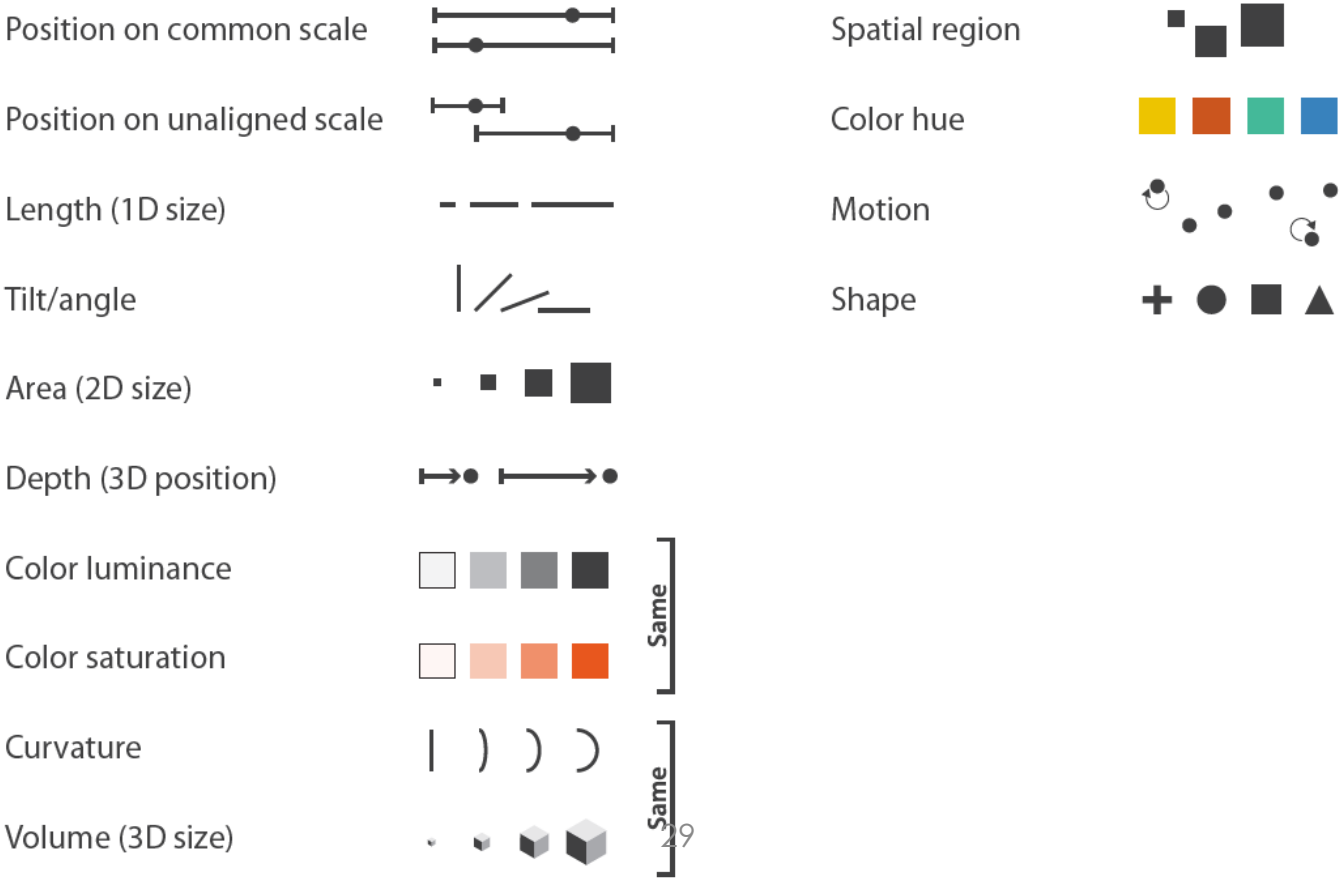
3:
vertical position
horizontal position
color hue

mark: point

4:
vertical position
horizontal position
color hue
size (area)

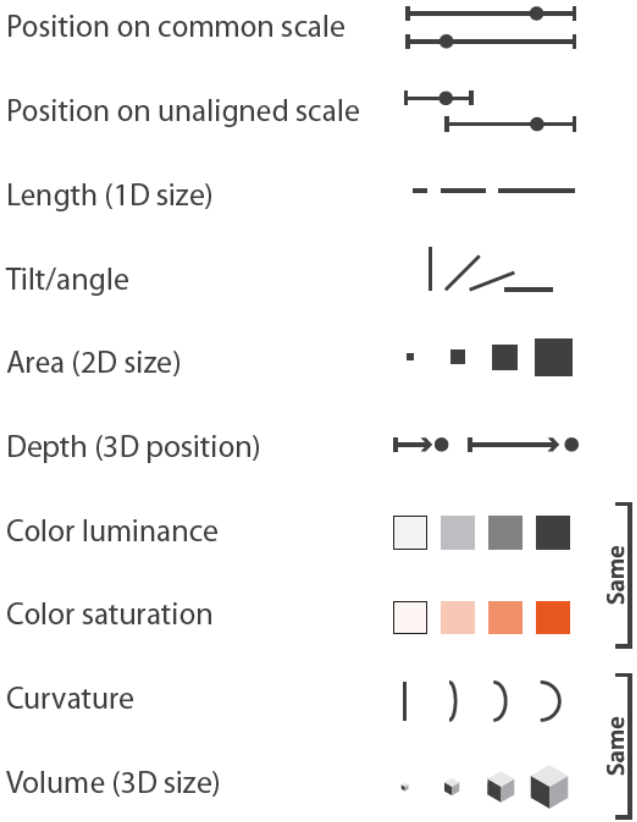
mark: point

Channels

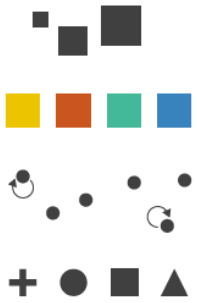

Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Same
Same
Spatial region
Color hue
Motion
Shape

Channels: Matching Types

Magnitude Channels: Ordered Attributes



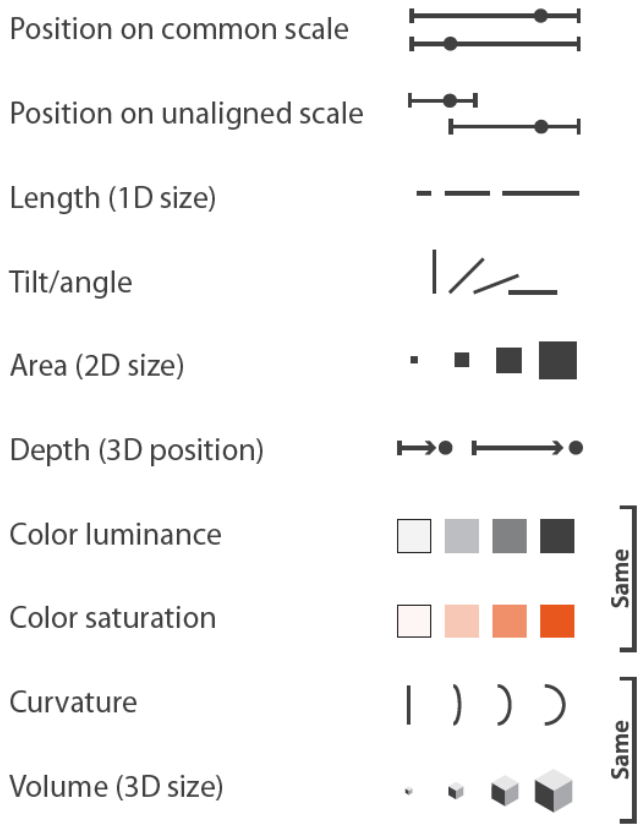
Identity Channels: Categorical Attributes



- expressiveness principle
  - match channel and data characteristics

Channels: Rankings

Magnitude Channels: Ordered Attributes



Identity Channels: Categorical Attributes

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

Questions?

Introduction to Visualization

part 4

Noeska Smit, Jan Byška *et al.*,
UiB Dept. of Informatics,
2017-08-28

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

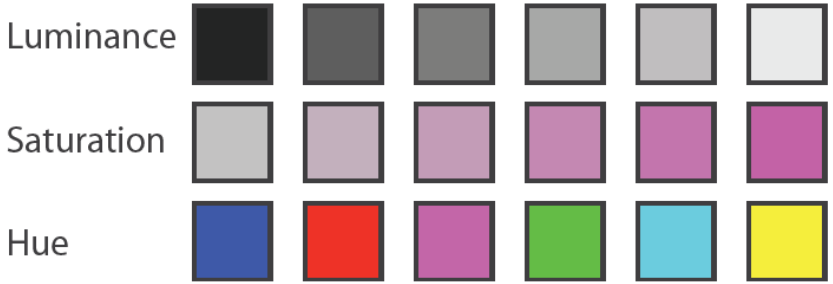
- Motion

  Direction, Rate, Frequency, ...

Color

Color can be decomposed into three channels:

- Ordered can show magnitude:
  - Luminance
  - Saturation
- Categorical can show identity:
  - Hue

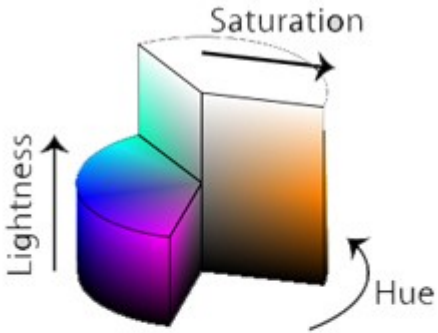


Color Systems

CMYK - Subtractive Color

RGB - Additive Color

HSL

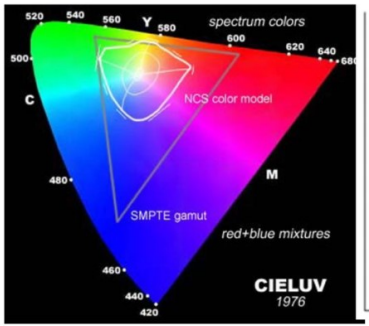


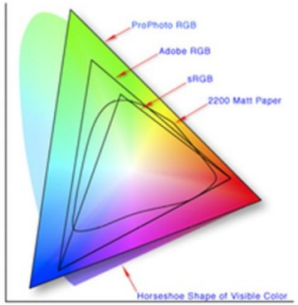
UNC

CIE updated:

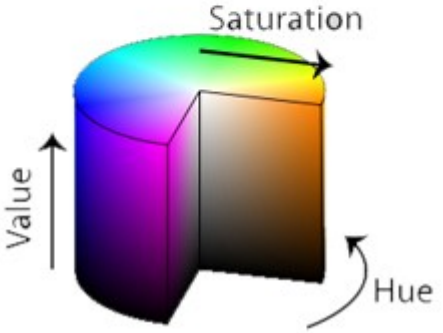
CIELAB (print)

CIELUV (display)



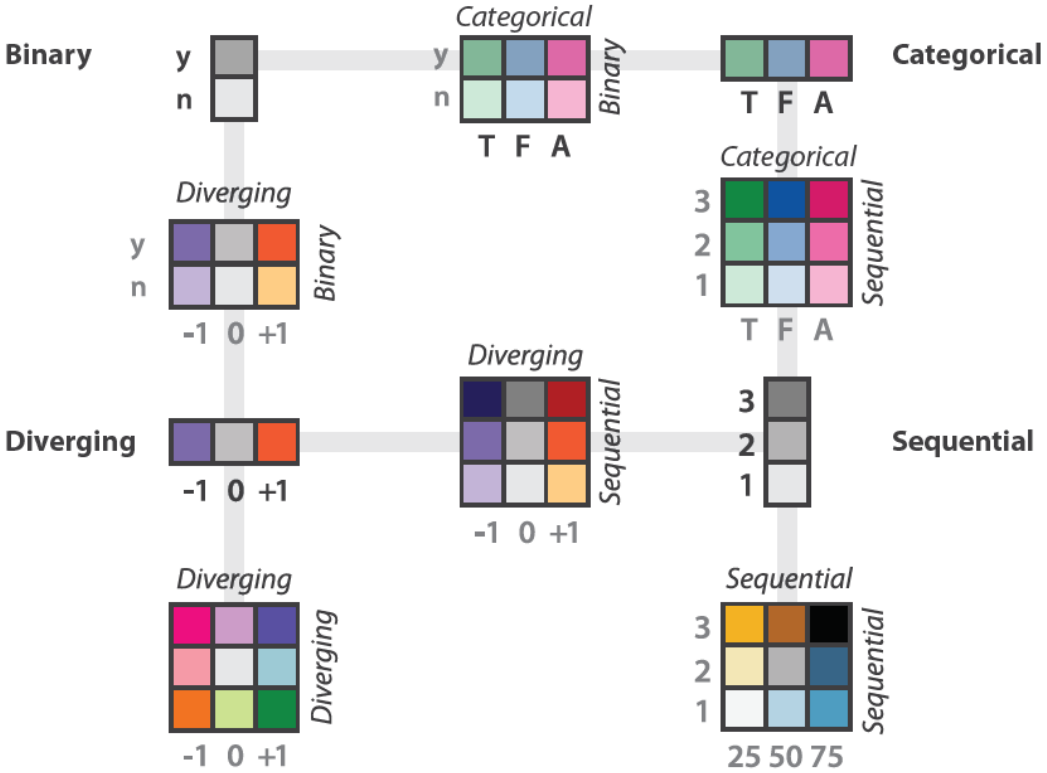


HSV



Color Maps <-> Data Types

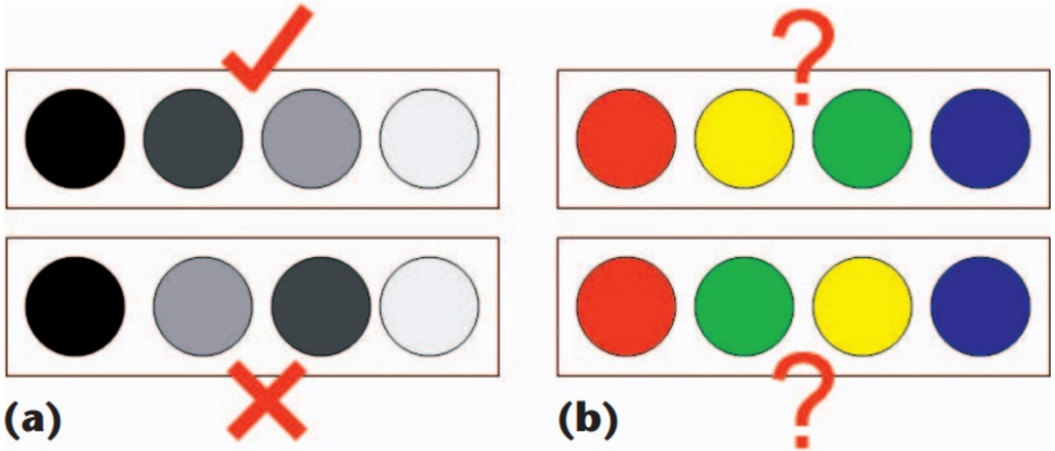
- Categorical
- Ordered
  - *Sequential*
  - *Diverging*
- Bivariate

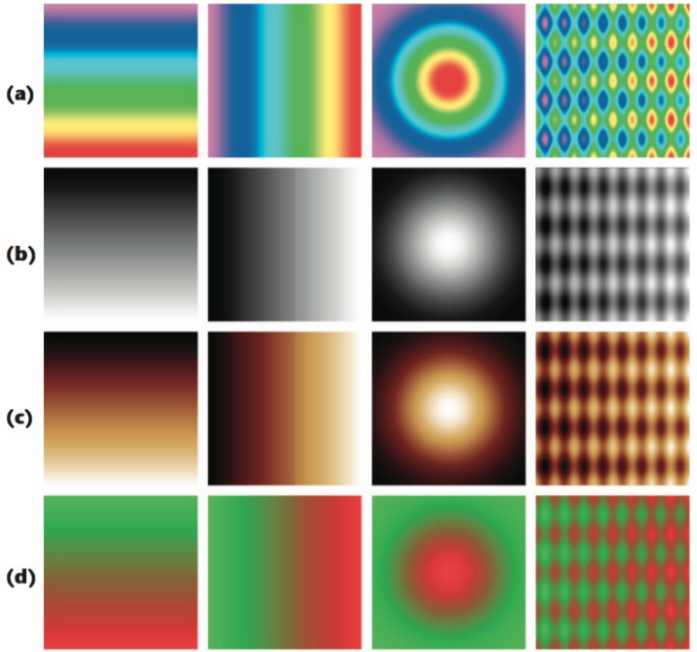


after [Color Use Guidelines for Mapping and Visualization. Brewer, 1994. http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html]

The Rainbow (Jet) Colormap

No perceptual ordering

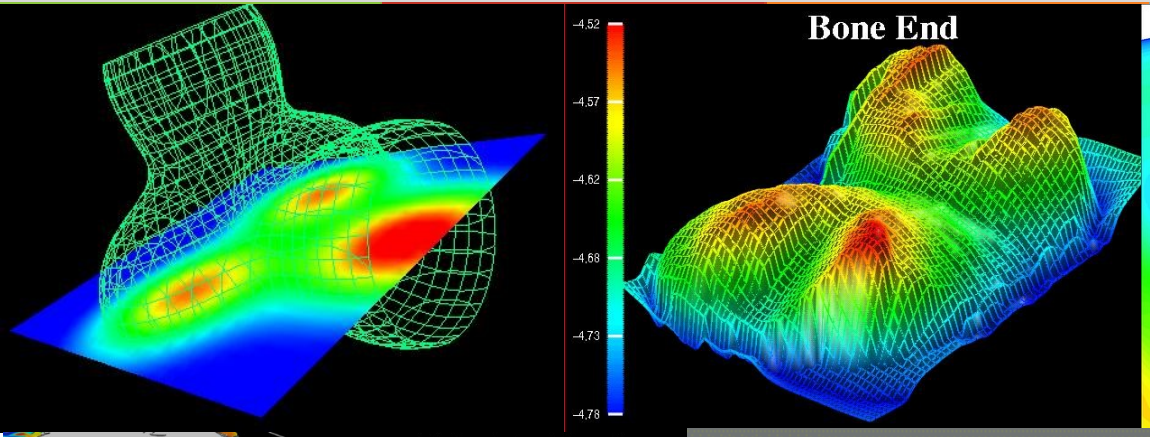




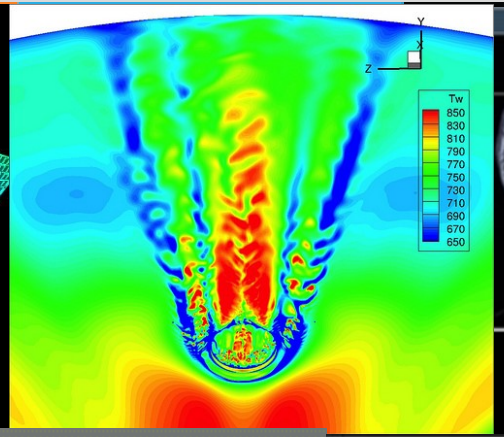
Not perceptually linear

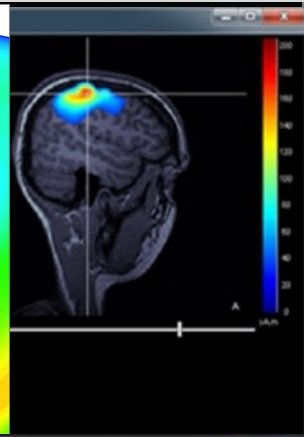
http://medvis.org/2012/08/21/rainbow-colormaps-what-are-they-good-for-absolutely-nothing/

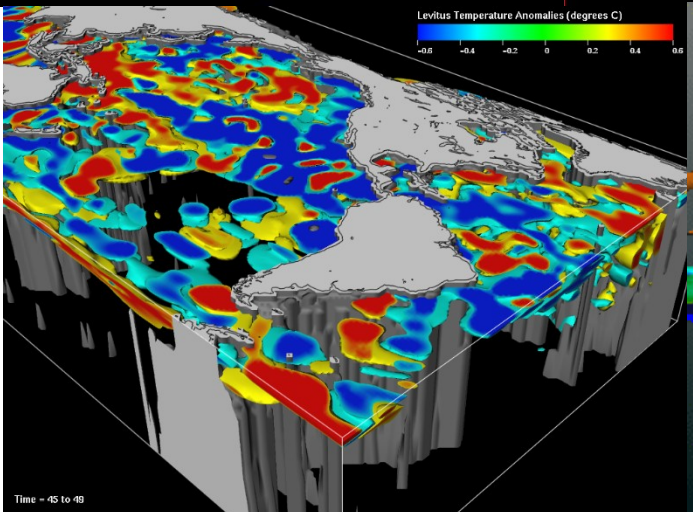
And yet... Jet!

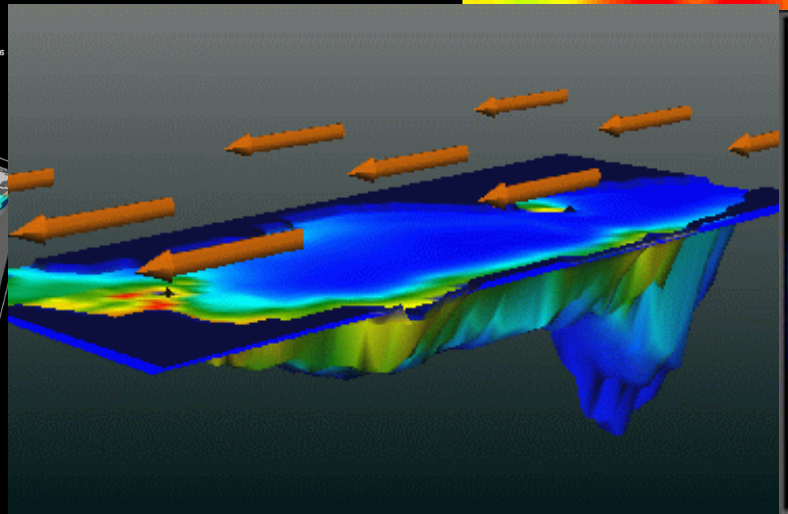

-4.52
-4.57
-4.62
-4.68
-4.73
-4.78

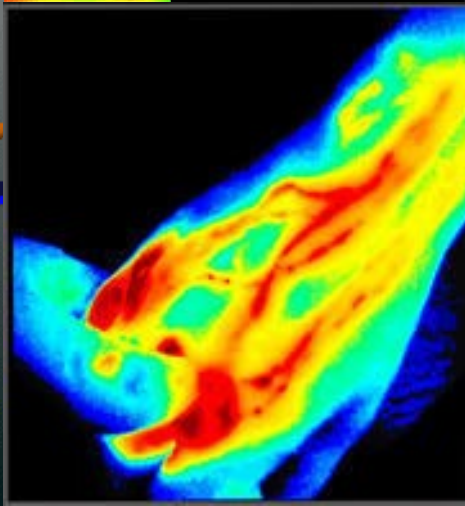
Bone End


Y
X
Z
Tw
850
830
810
790
770
750
730
710
690
670
650

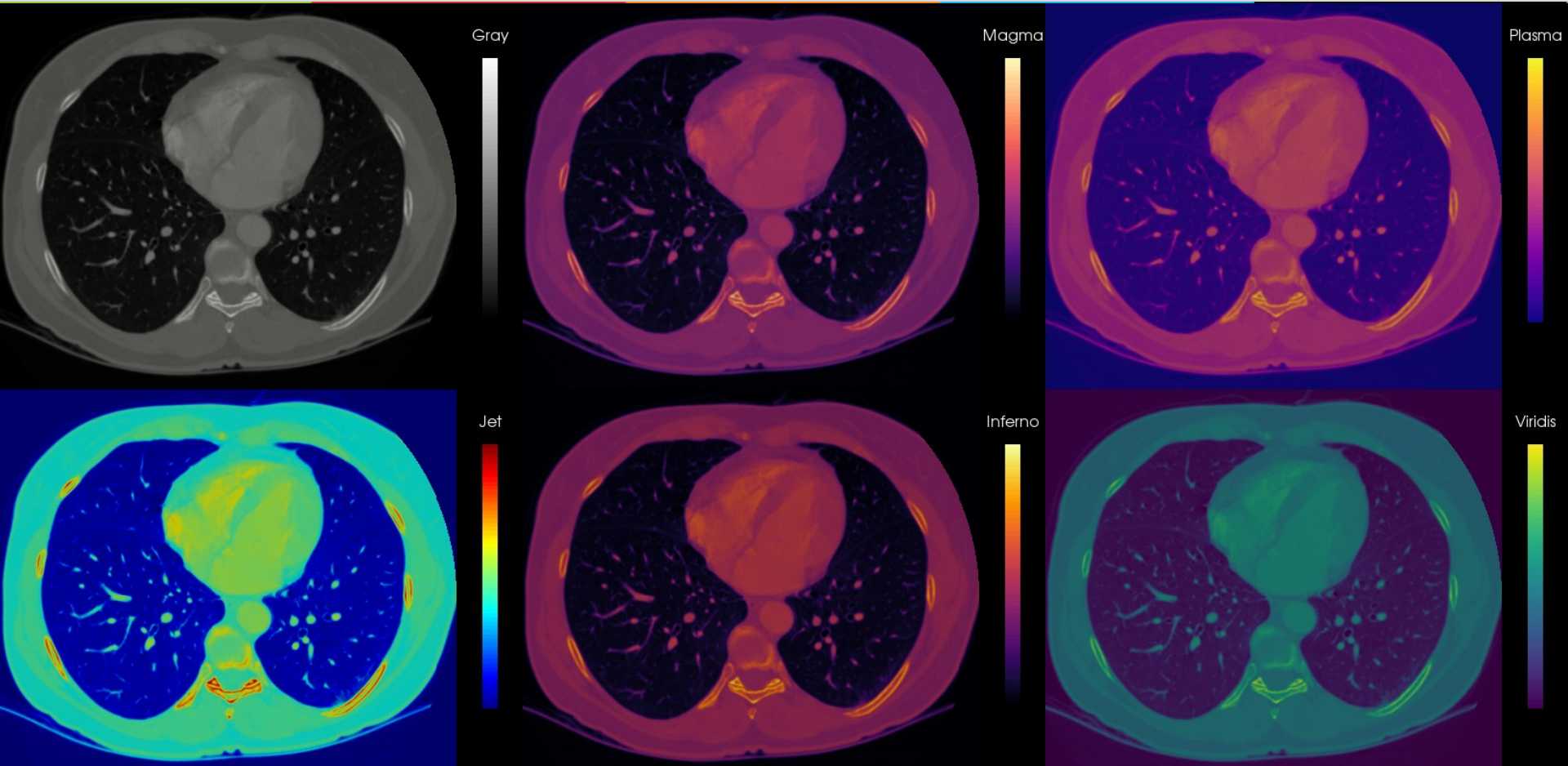



Levitus Temperature Anomalies (degrees C)
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
Time = 45 to 49





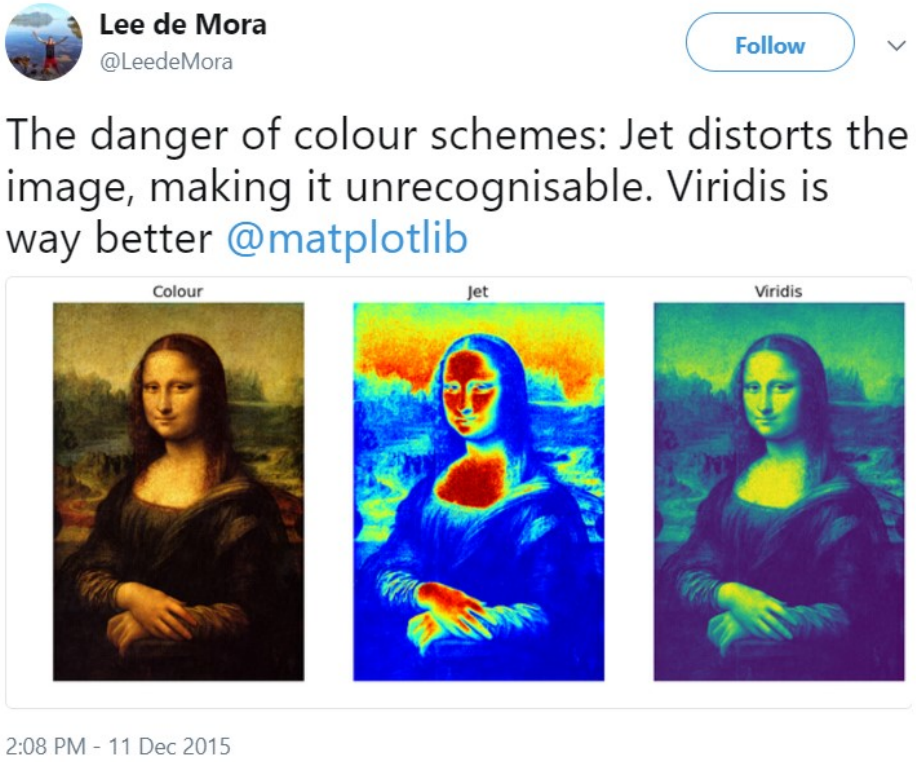
The rainbow (Jet) vs. Matplotlib's new collection



Viridis, Magma, Plasma, and Inferno

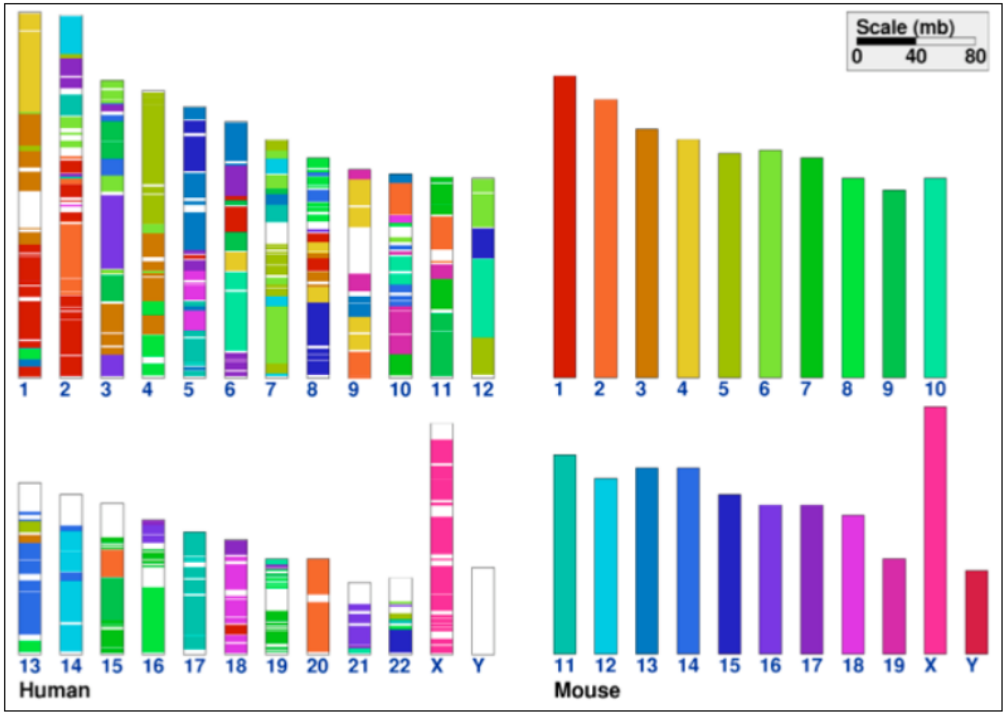
https://matplotlib.org/users/colormaps.html

- monotonically increasing luminance, perceptually uniform
- colorful, colourblind-safe
- R, Python, D3
- In Matlab: Parula (new default) is close to Viridis



Categorical Color: Discriminability constraints

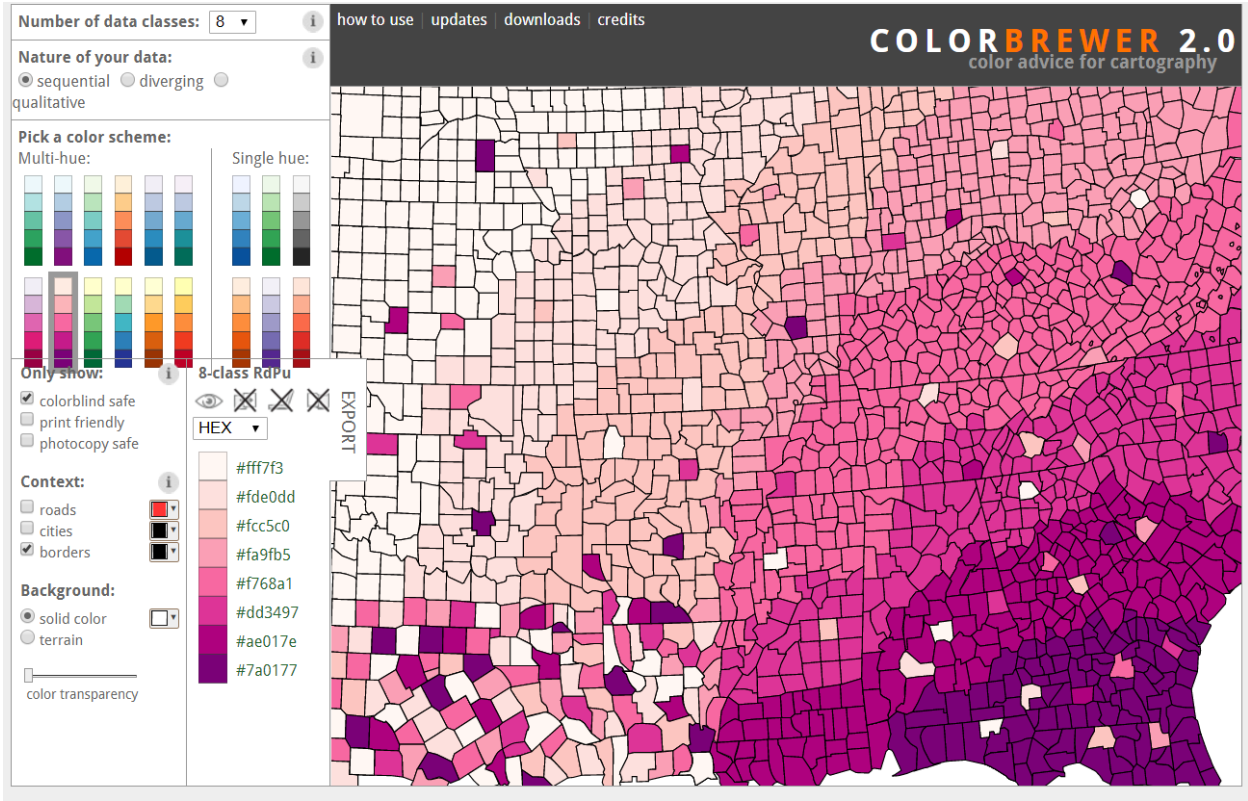
- noncontiguous small regions of color: only 6-12 bins



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

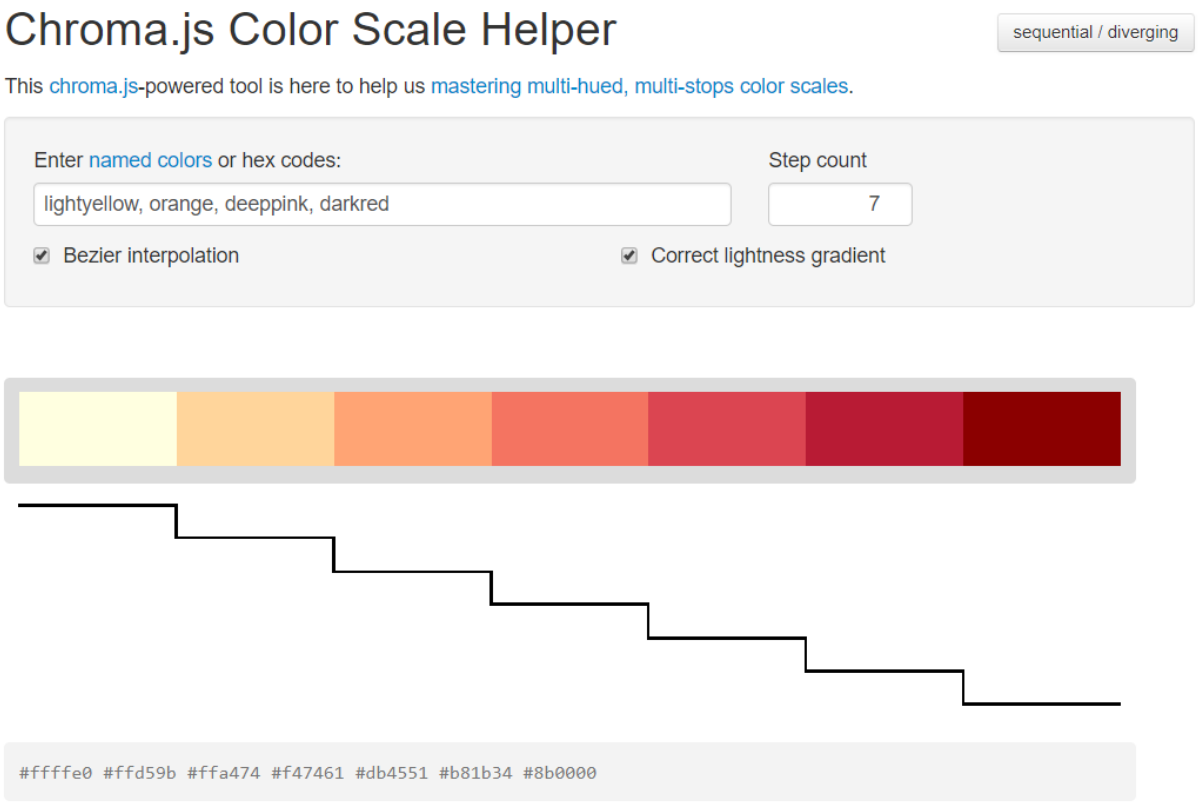
Useful Color Tools (1) – Colors for Maps

http://colorbrewer2.org/



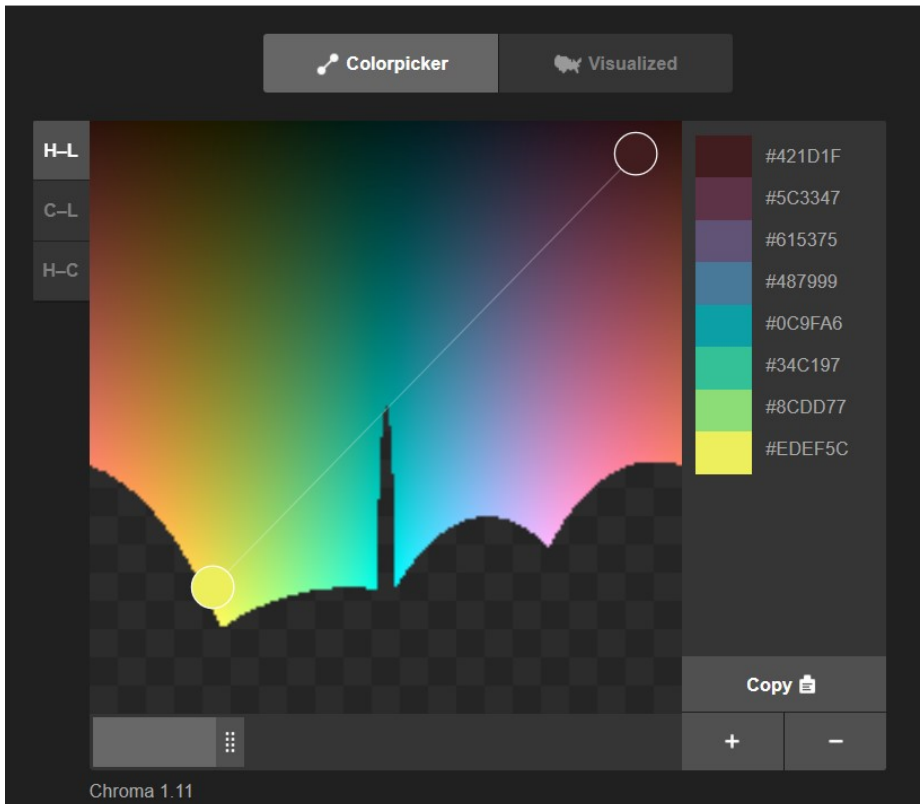
Useful Color Tools (2) - Color Scales

http://gka.github.io/palettes/



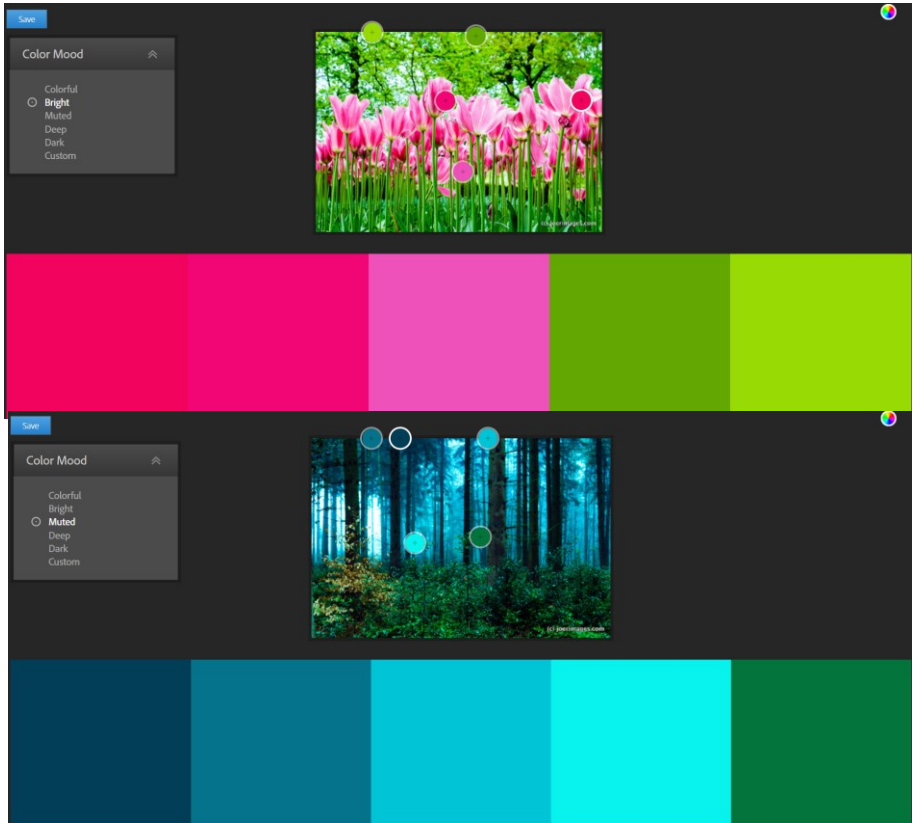
Useful Color Tools (3) – Perceptually Linear

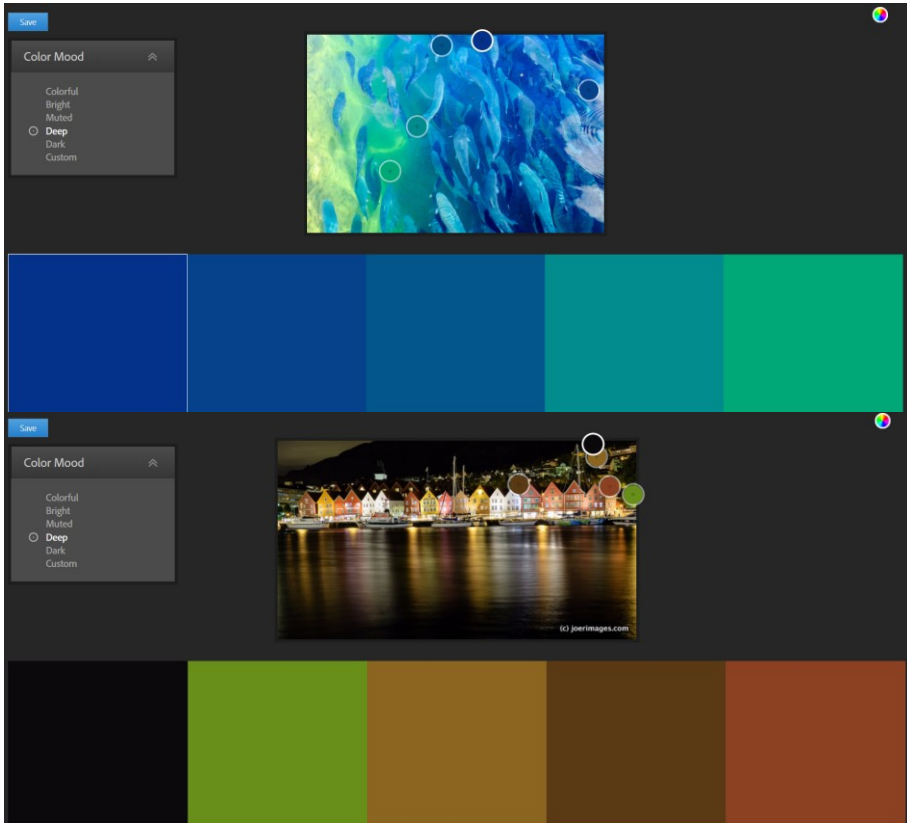
http://tristen.ca/hcl-picker/



Fun Color Tool – Adobe Color

https://color.adobe.com/create/image/ (+ photos by Joeri Smit)





How to handle data complexity?

How?

Manipulate	Facet	Reduce
Change 	Juxtapose 	Filter 
Select 	Partition 	Aggregate 
Navigate 	Superimpose 	Embed 

Handling complexity:

Manipulate	Facet	Reduce	→ Derive
→ Change 	→ Juxtapose 	→ Filter 	
→ Select 	→ Partition 	→ Aggregate 	
→ Navigate 	→ Superimpose 	→ Embed 	

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

28

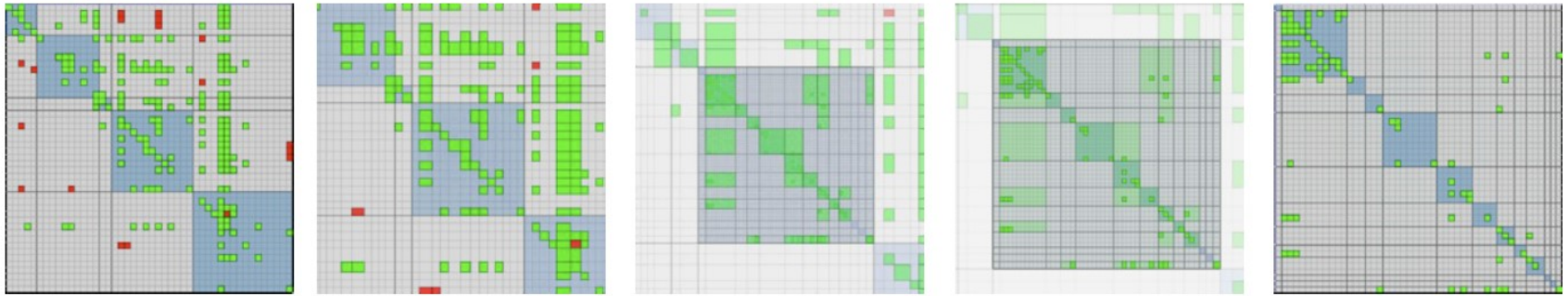
Idiom: Animated Transitions

- Smooth transition from one state to another
- Can track when change is limited

Example: multilevel matrix views:

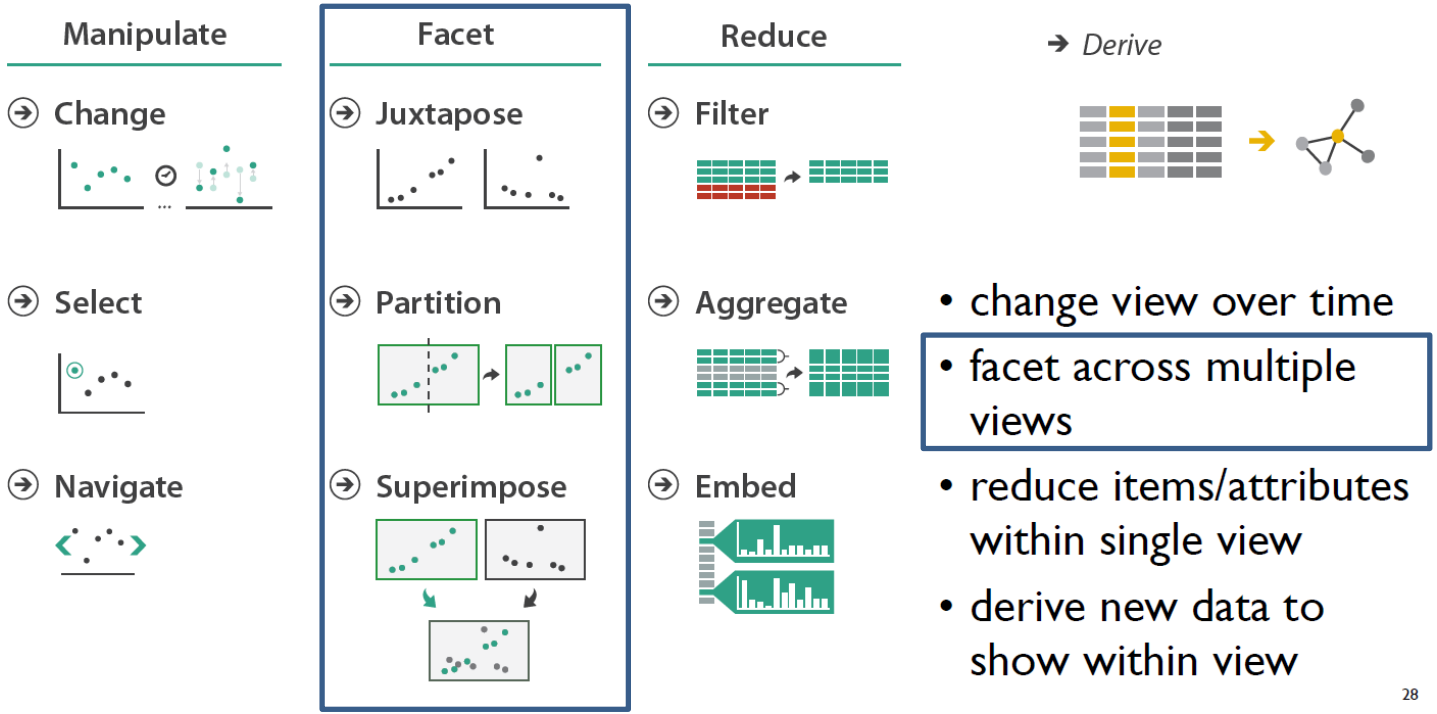
– scope of what is shown narrows down

- middle block stretches to fill space, additional structure appears within
- other blocks squish down to increasingly aggregated representations



[Using Multilevel Call Matrices in Large Software Projects. van Ham. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 227–232, 2003.]

Handling complexity:



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

28

Facet

Coordinate Multiple Side By Side Views

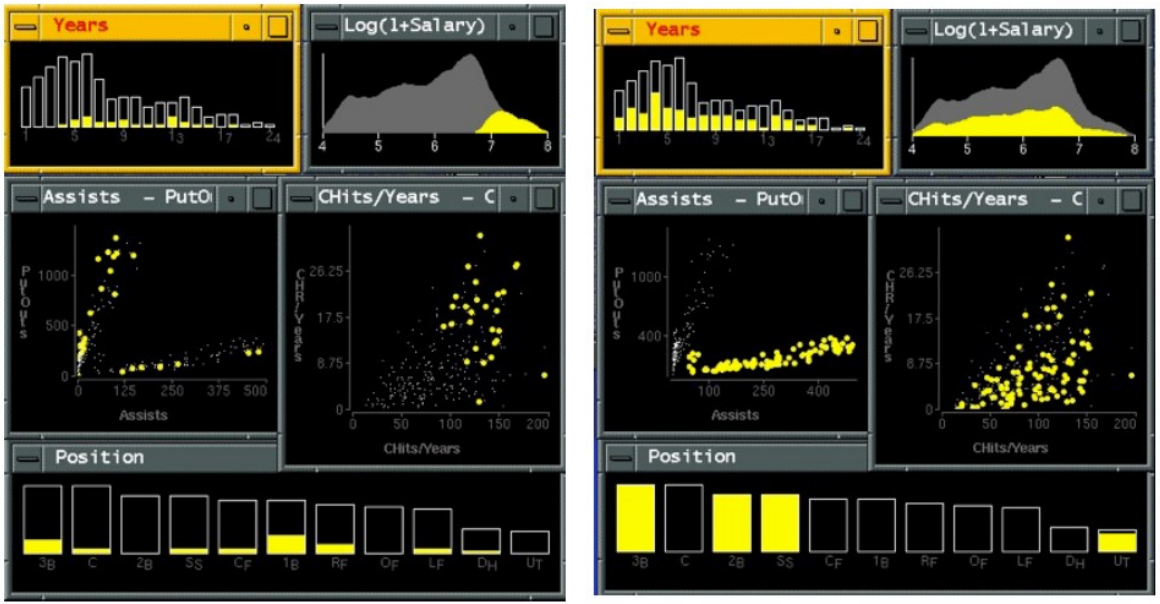
- Share Encoding: Same/Different
  - *Linked Highlighting*

- Share Data: All/Subset/None

- Share Navigation

Idiom: Linked Highlighting

see how regions in one view are distributed within another via interaction: encoding: different, data: all shared



[Visual Exploration of Large Structured Datasets. Wills. Proc. New Techniques and Trends in Statistics (NTTS), pp. 237–246. IOS Press, 1995.]

Idiom: Bird's Eye Maps

encoding: same

data: subset shared

navigation: shared

differences: viewpoint and size

[A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.]

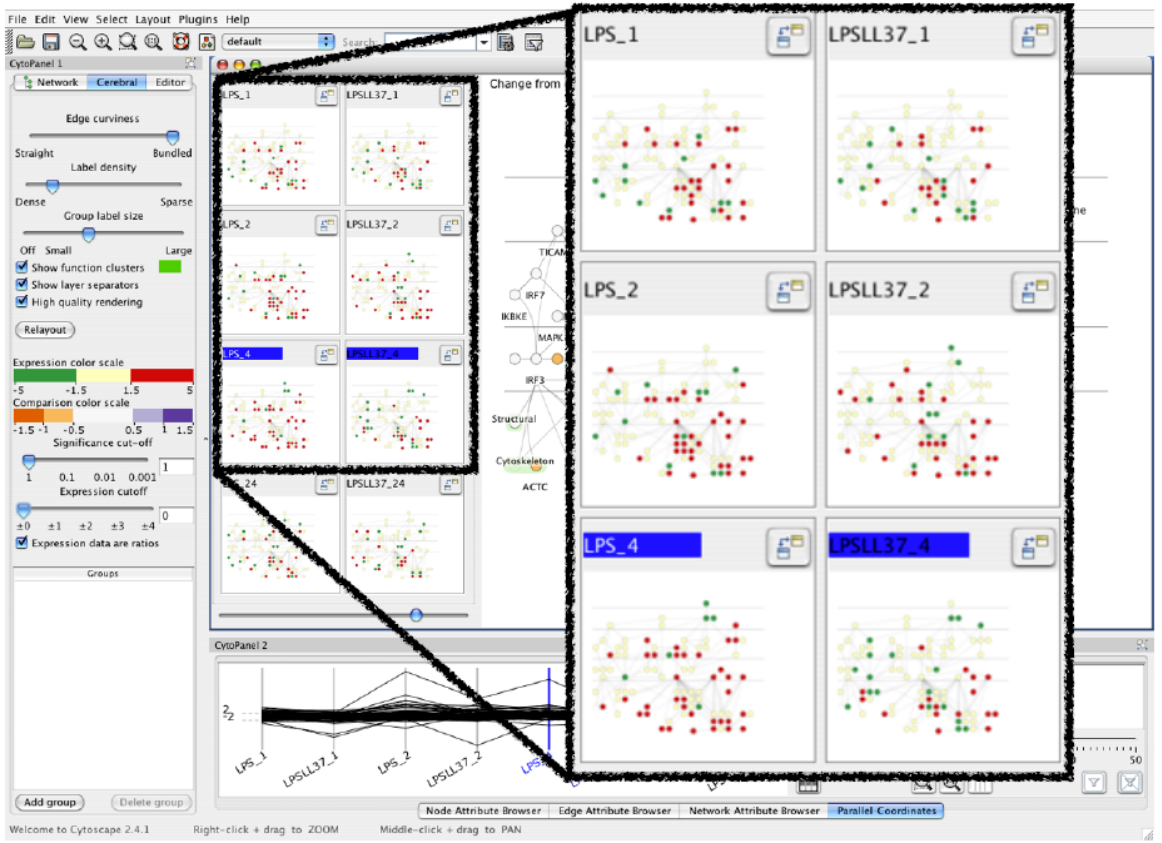
overview-detail

Idiom: Small Multiples

encoding: same

data: none shared (different attributes for colors)

navigation: shared

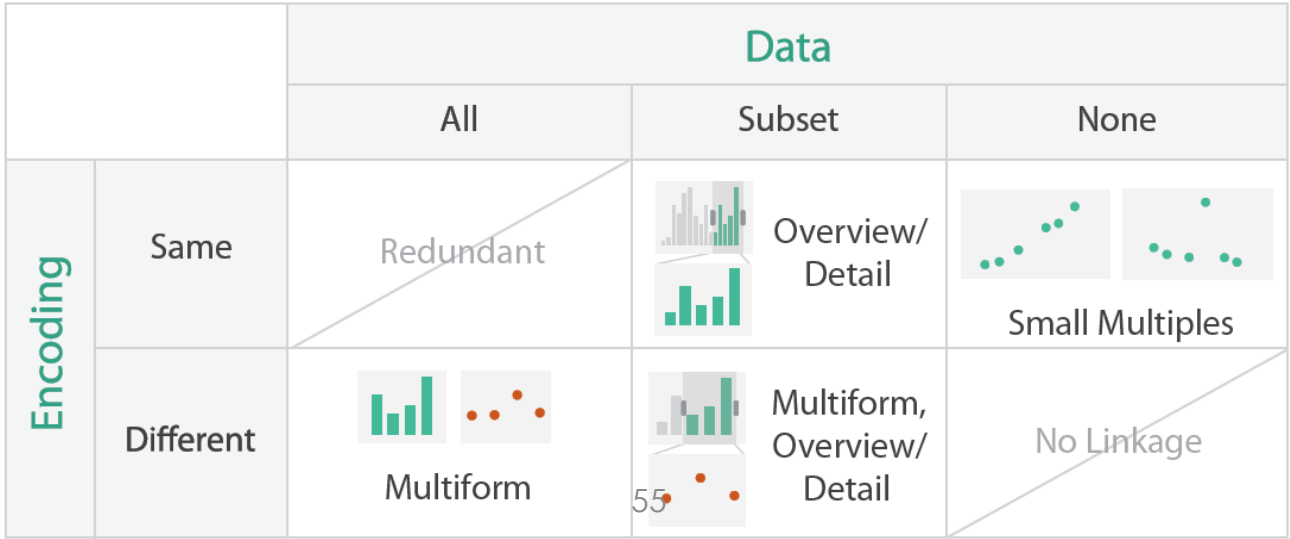


[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG 2008]

Coordinate views: Design choice interaction

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



Partition into views

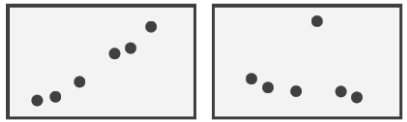
how to divide data between views

- encodes association between items using spatial proximity
- major implications for what patterns are visible
- split according to attributes

design choices

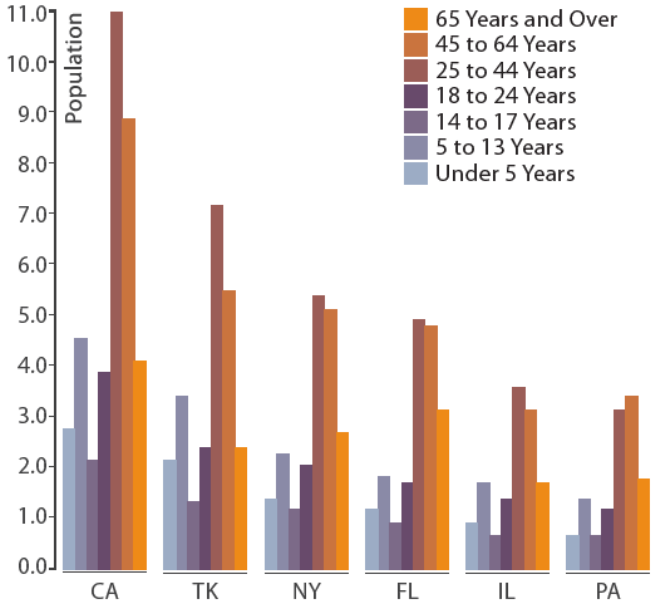
- how many splits
  - all the way down: one mark per region?
  - stop earlier, for more complex structure within region?
- order in which attributes used to split
- how many views

Partition into Side-by-Side Views



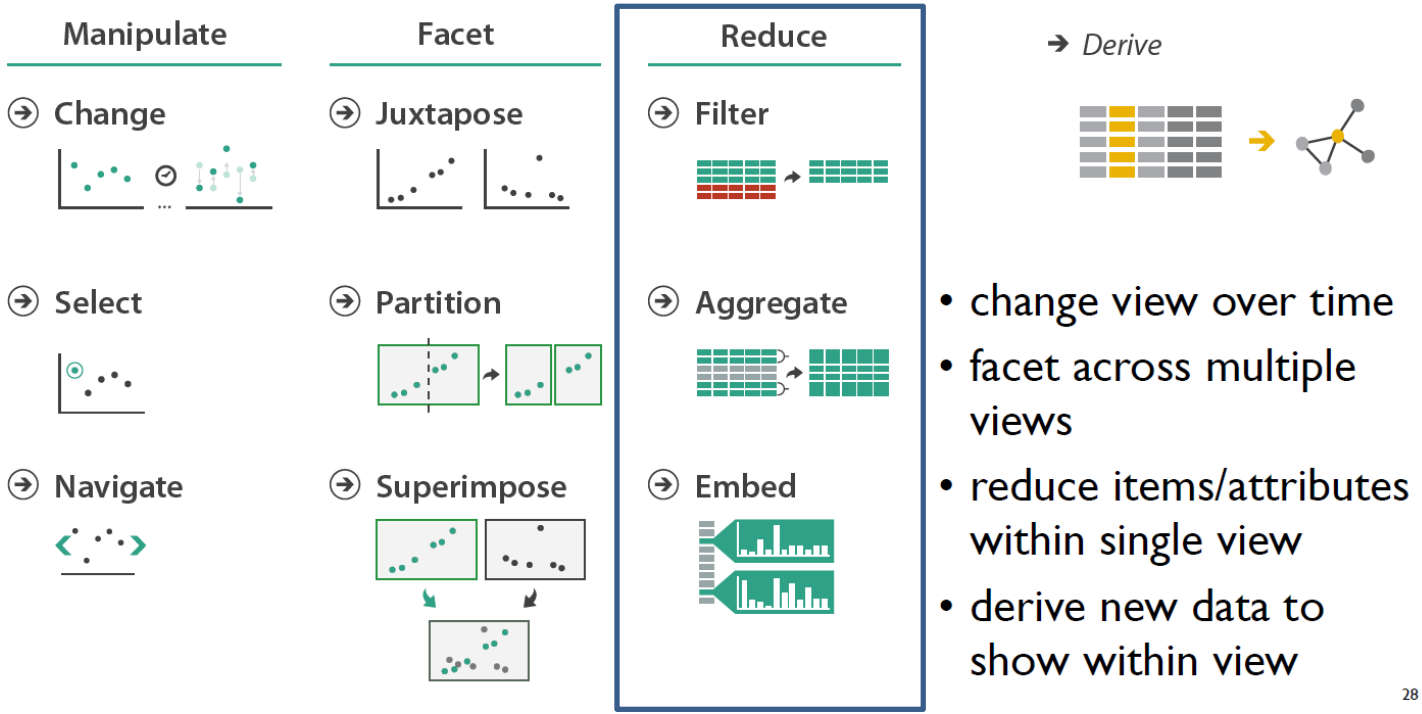
Partitioning: List alignment

- 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



http:/bl.ocks.org/mbostock/3887051 http:/bl.ocks.org/mbostock/4679202

Handling complexity:



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

28

Reduce items and attributes

reduce/increase: inverses

filter

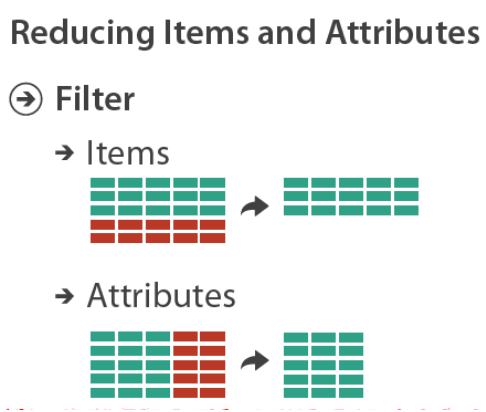
- pro: straightforward and intuitive to understand and compute
- con: out of sight, out of mind

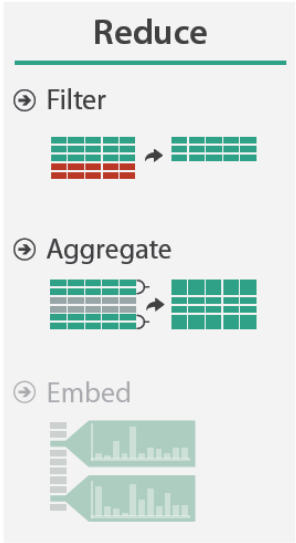
aggregation

- pro: inform about whole set
- con: difficult to avoid losing signal

not mutually exclusive

- combine filter, aggregate
- combine reduce, facet, change, derive





Idom: boxplot

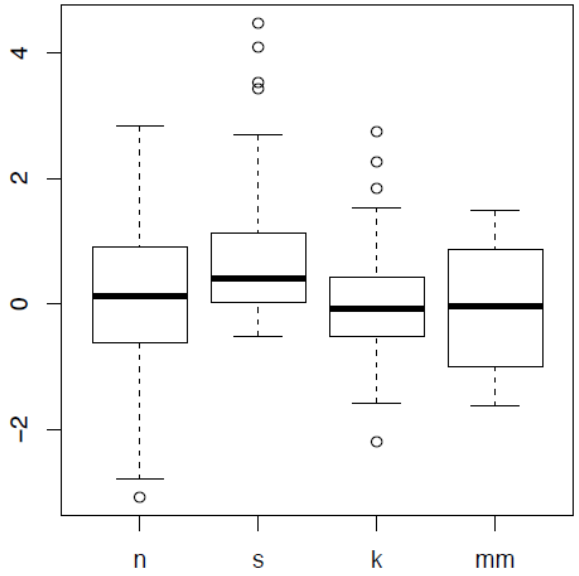
static item aggregation

task: find distribution

data: table

derived data

- 5 quant attribs
  - median: central line
  - lower and upper quartile: boxes
  - lower upper fences: whiskers
    - values beyond which items are outliers

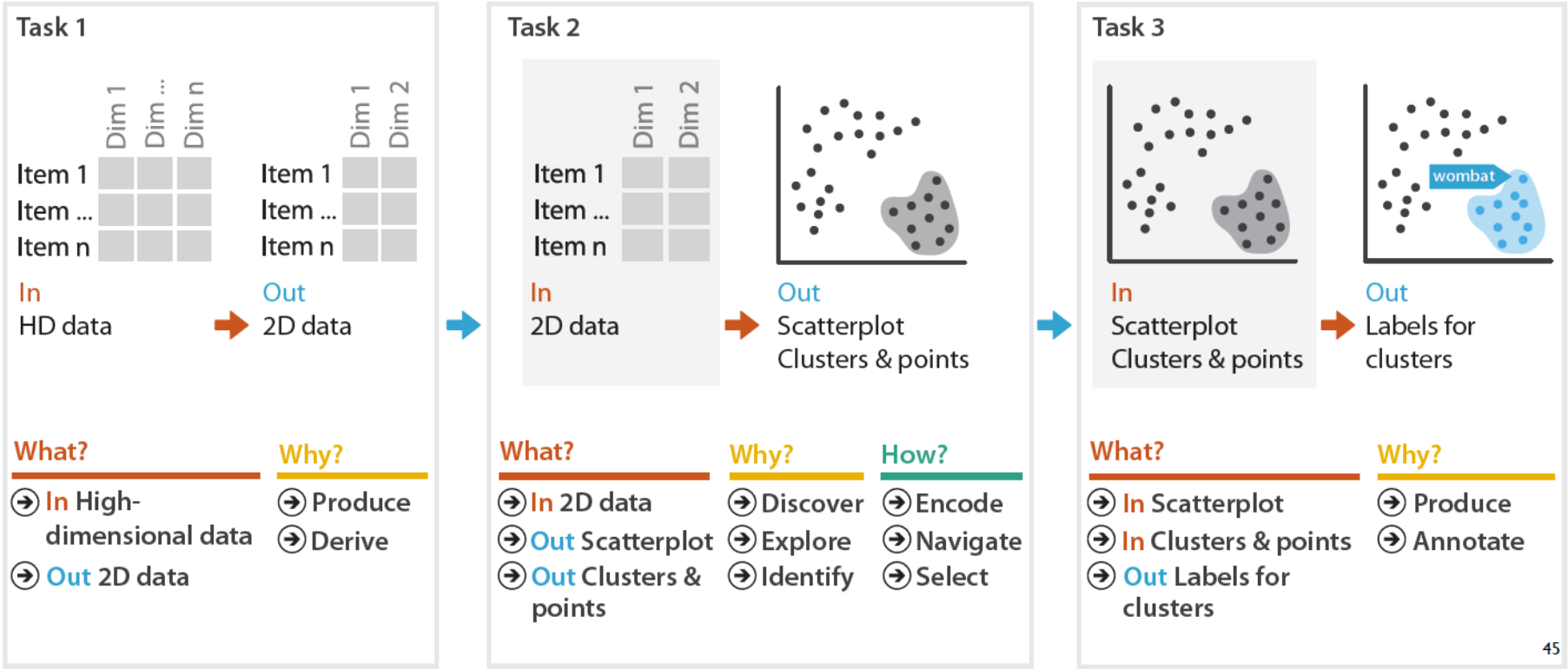


outliers beyond fence cutoffs explicitly shown

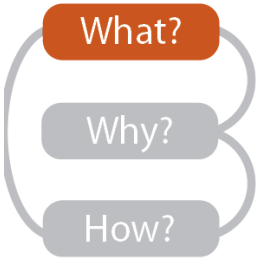
[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

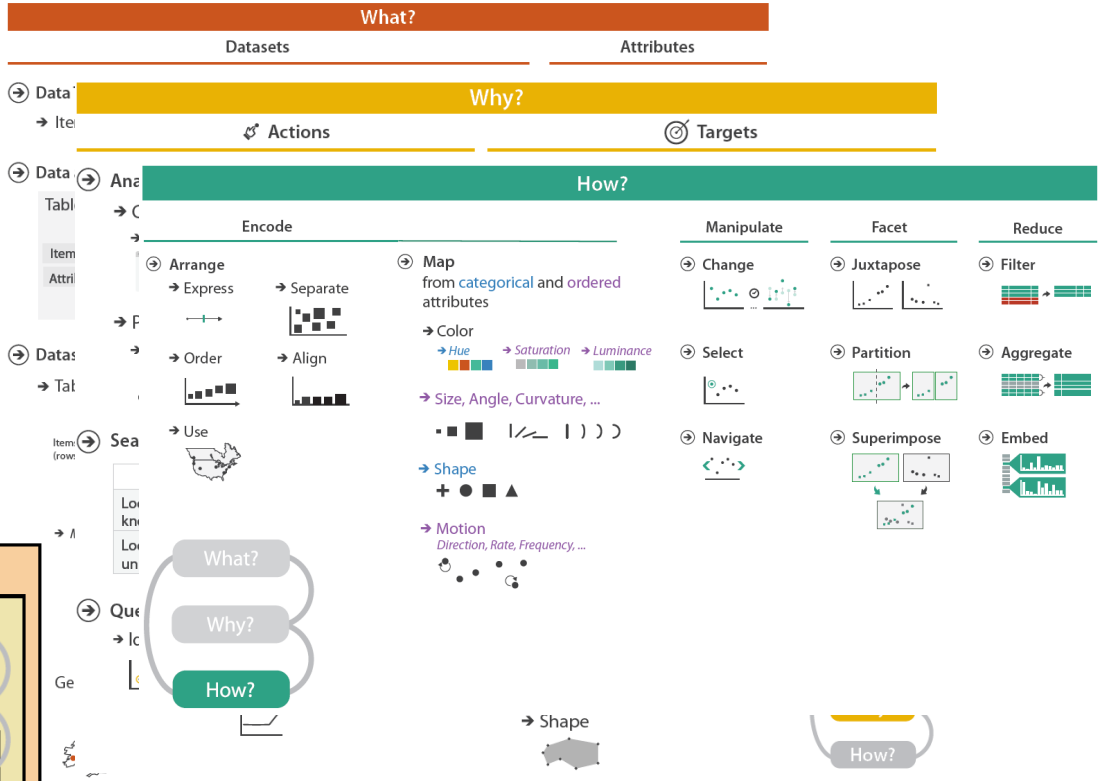
Idiom: Dimensionality reduction

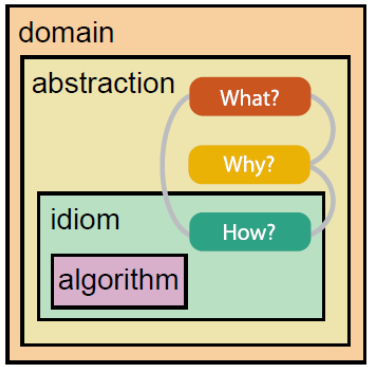
- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space



Summary


What?
Why?
How?


What?
Datasets
Attributes
Why?
Actions
Targets
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?
Shape
How?


domain
abstraction
What?
Why?
idiom
How?
algorithm

Acknowledgement

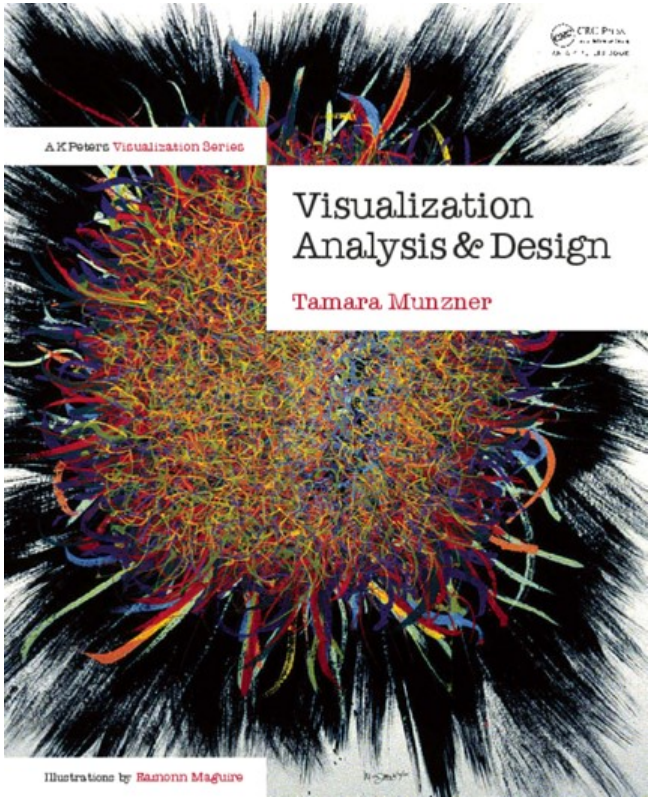
the source for this talk

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

book page (including tutorial lecture slides)

- http://www.cs.ubc.ca/~tmm/vadbook
- http://www.crcpress.com/product/isbn/9781466508910
- 20% promo code for book+ebook combo: HVN17

illustrations: Eamonn Maguirethe



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

Questions?