Information Theory and Visualization

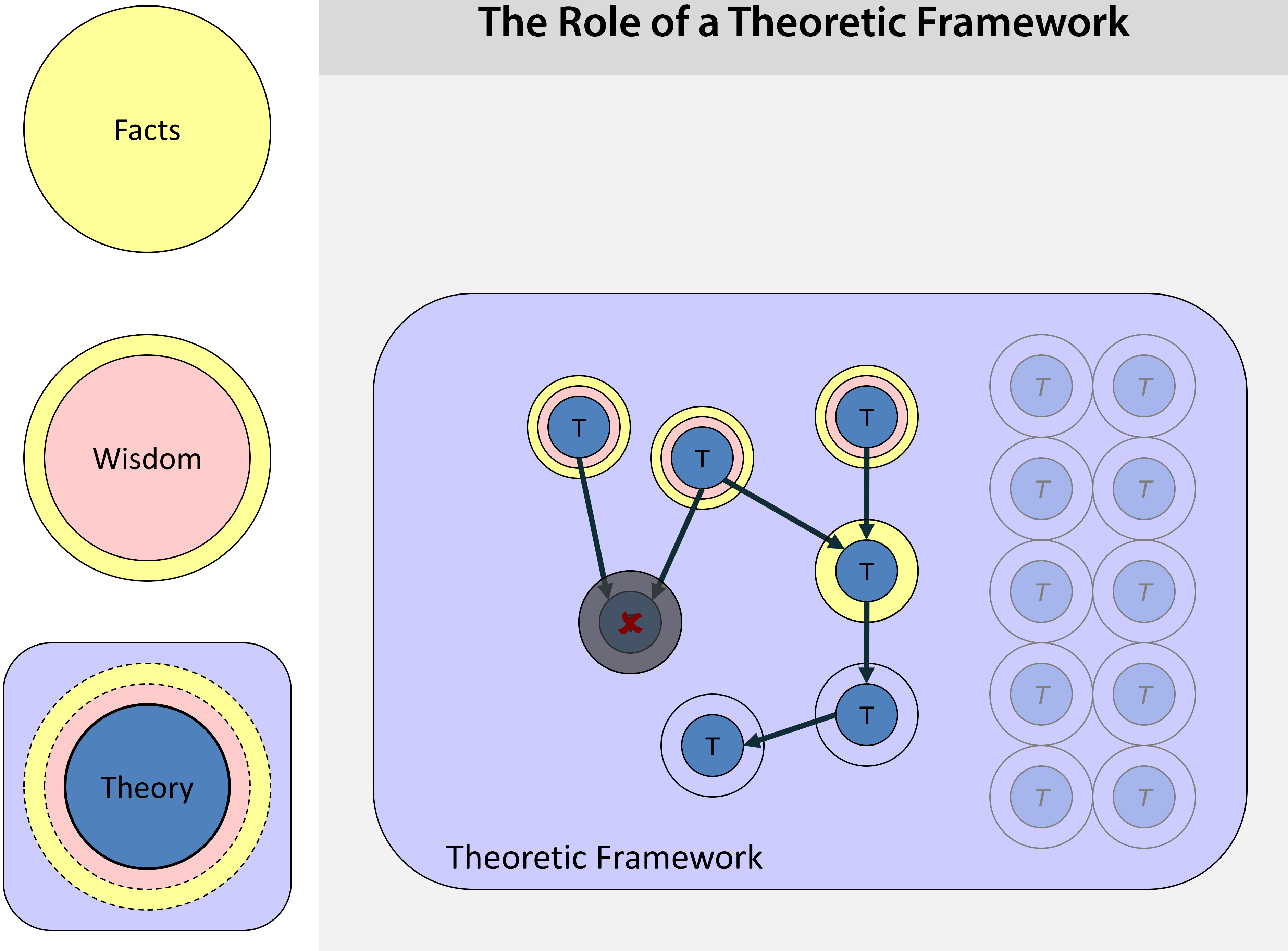
Tutorial on Information Theory in Visualization

Min Chen

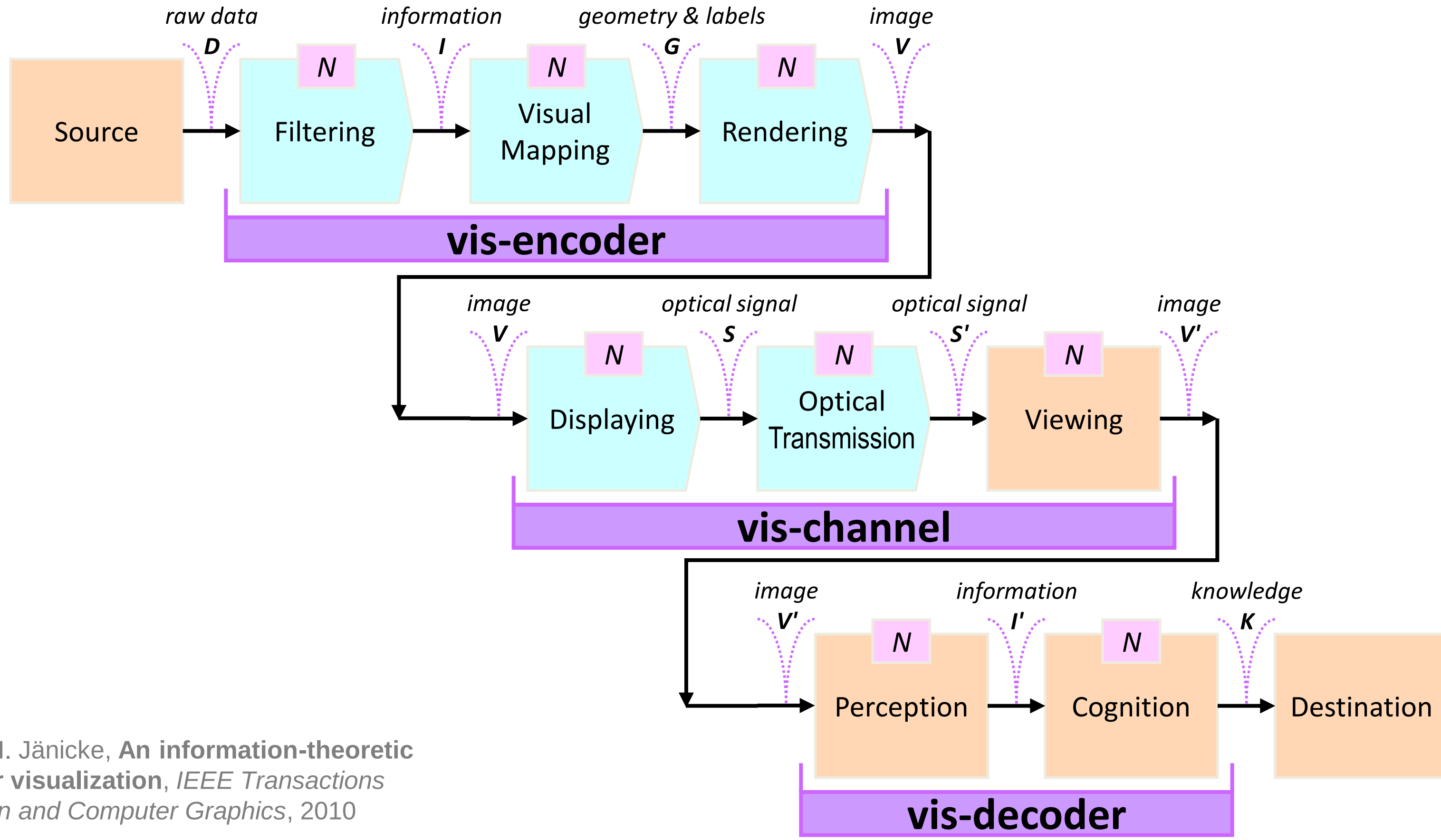
University of Oxford

EUROGRAPHICS 2016

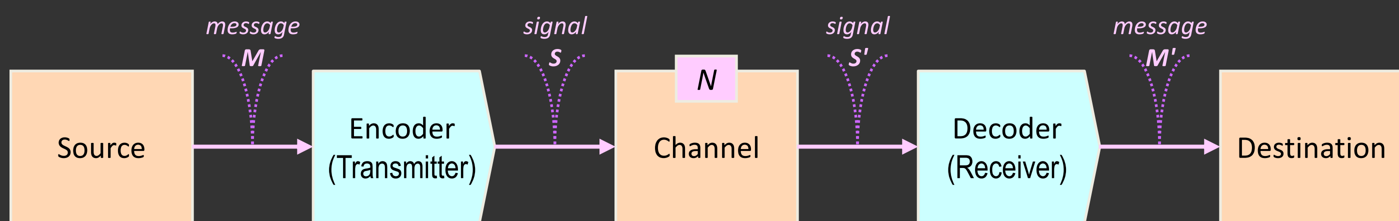
The Role of a Theoretic Framework



An Information-Theoretic View of Visualization

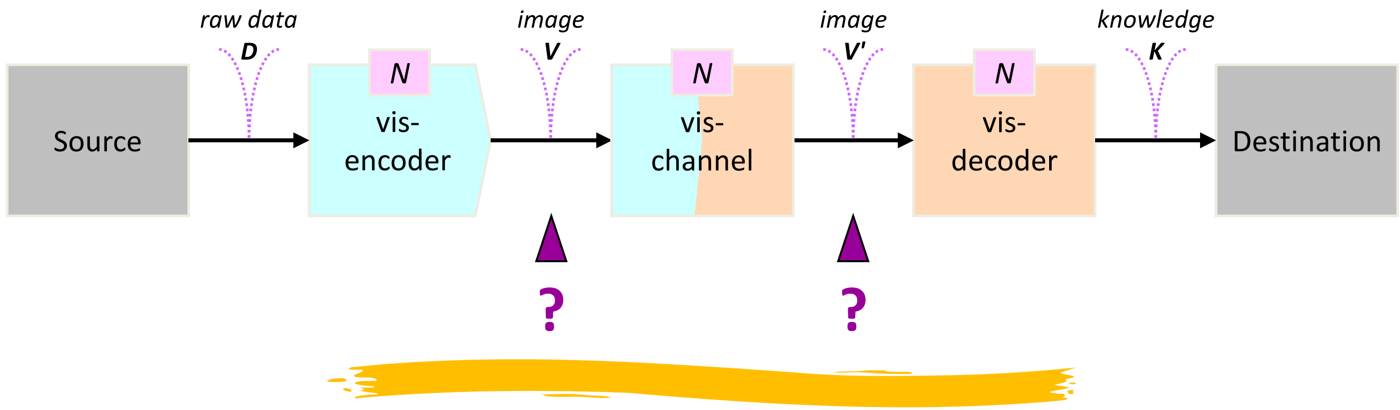


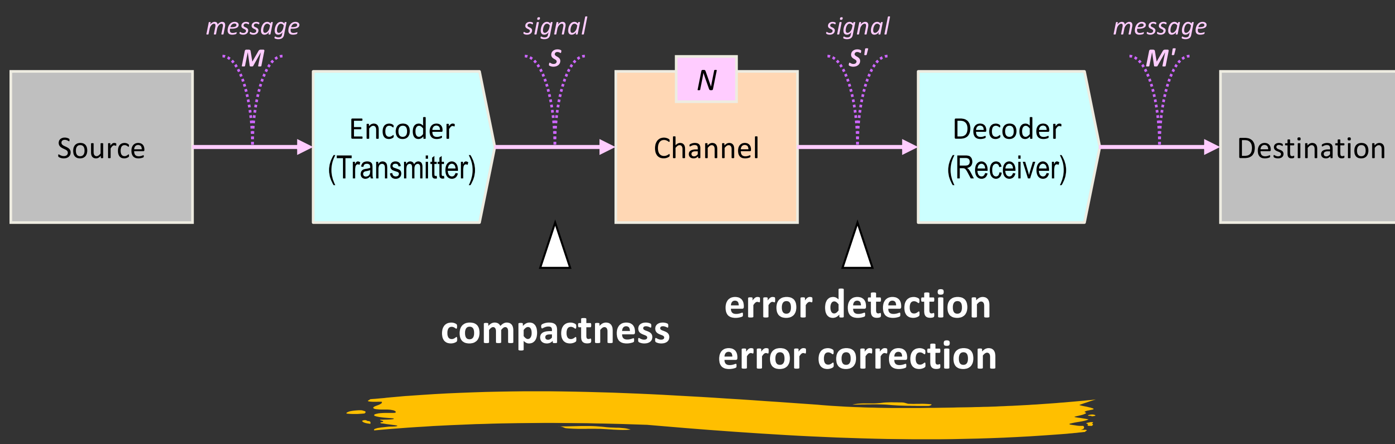
M. Chen and H. Jänicke, **An information-theoretic framework for visualization**, *IEEE Transactions on Visualisation and Computer Graphics*, 2010



A General Communication System

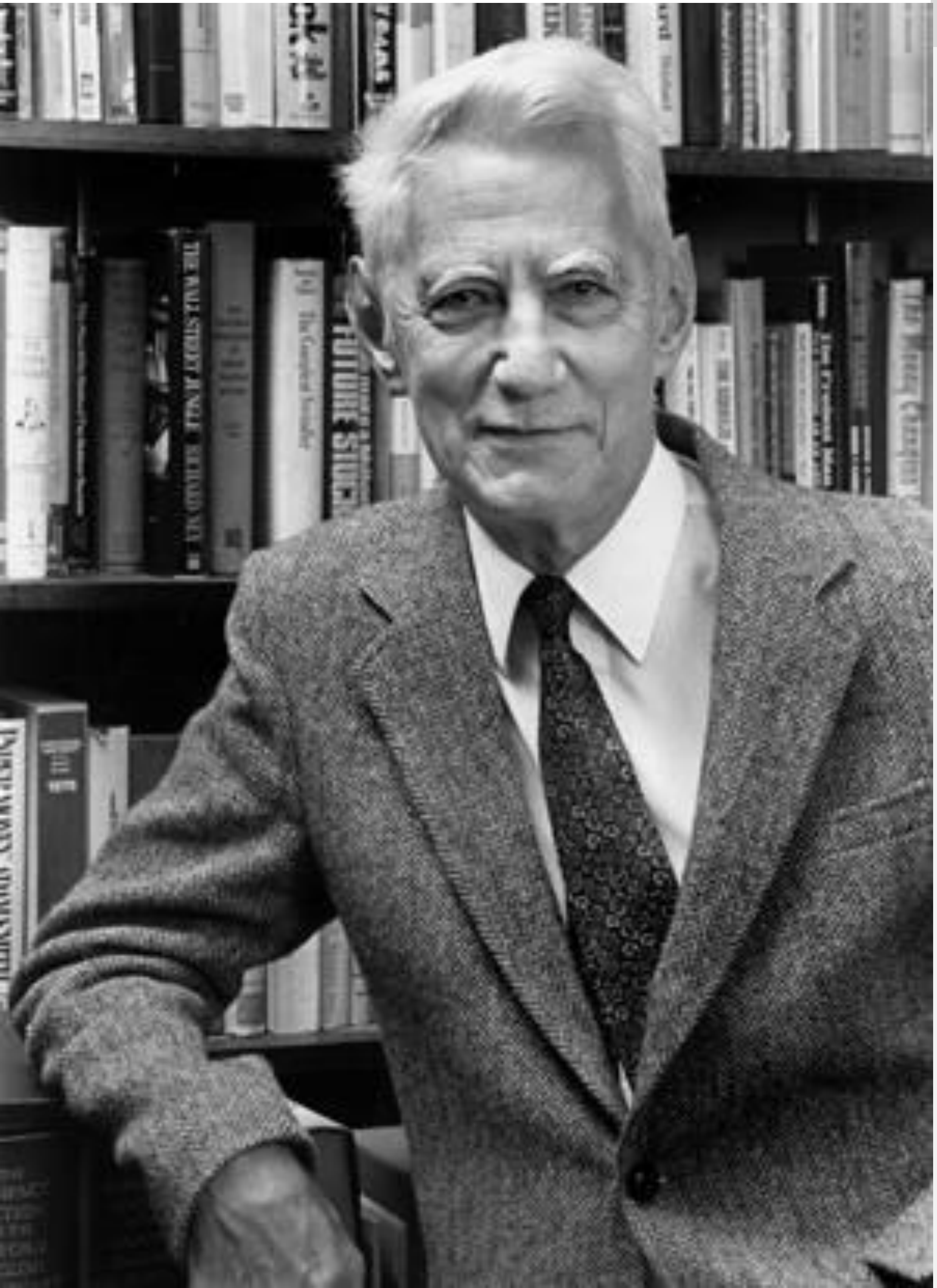
Three Visualization Subsystems





Three Communication Subsystems

Claude E. Shannon (1916-2001)



Entropy

- Random variable (alphabet)
  - X
- It takes values (letters)
  - $x_1, x_2, \ldots, x_m$
- Probability mass function
  - $p(x_i)$
- Entropy (uncertainty)

$$\mathbf{H}(X) = -\sum_{i}^{m} p(x_i) \log_2 p(x_i)$$

Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

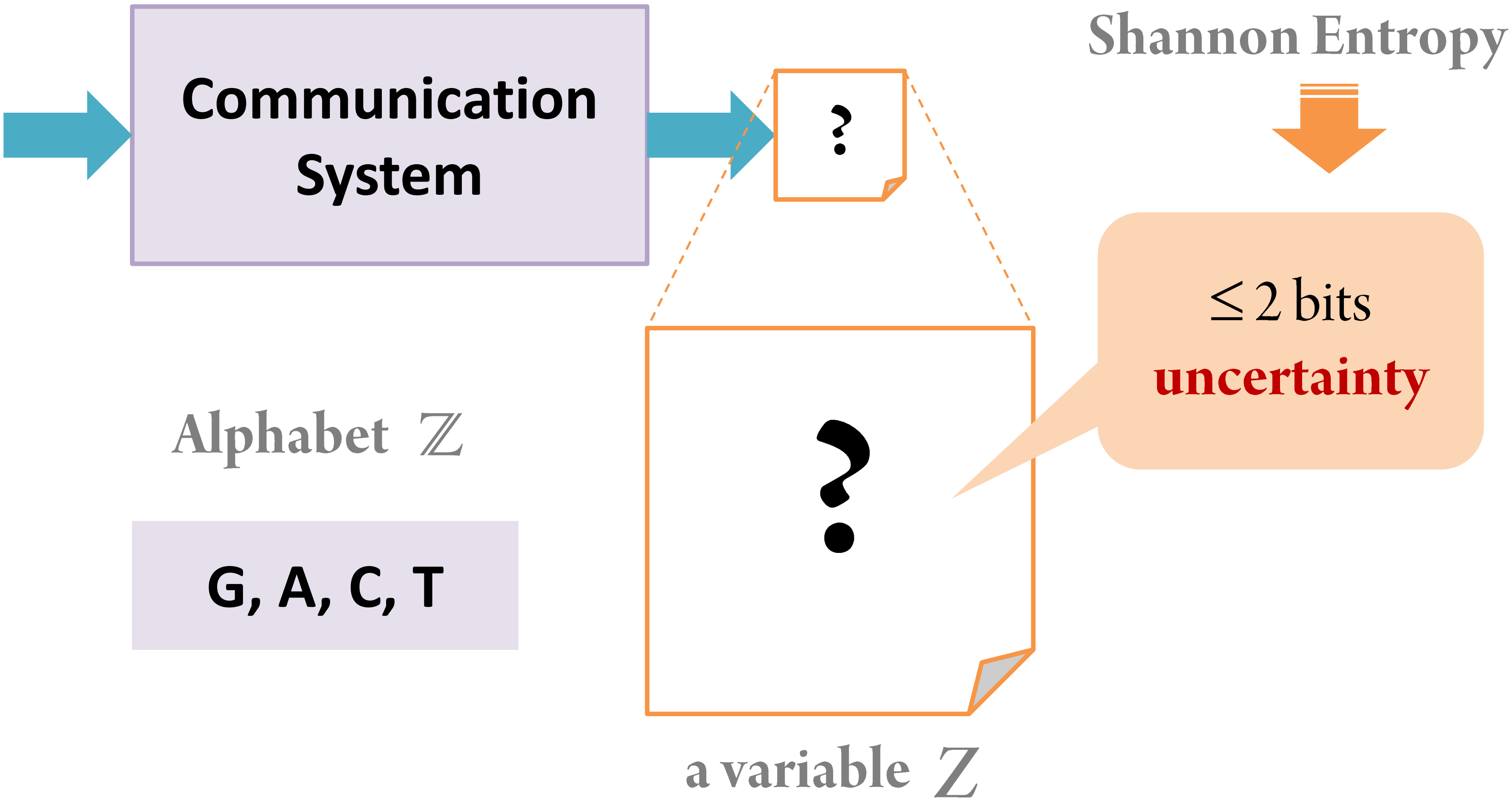
Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

Shannon's Definition of Information

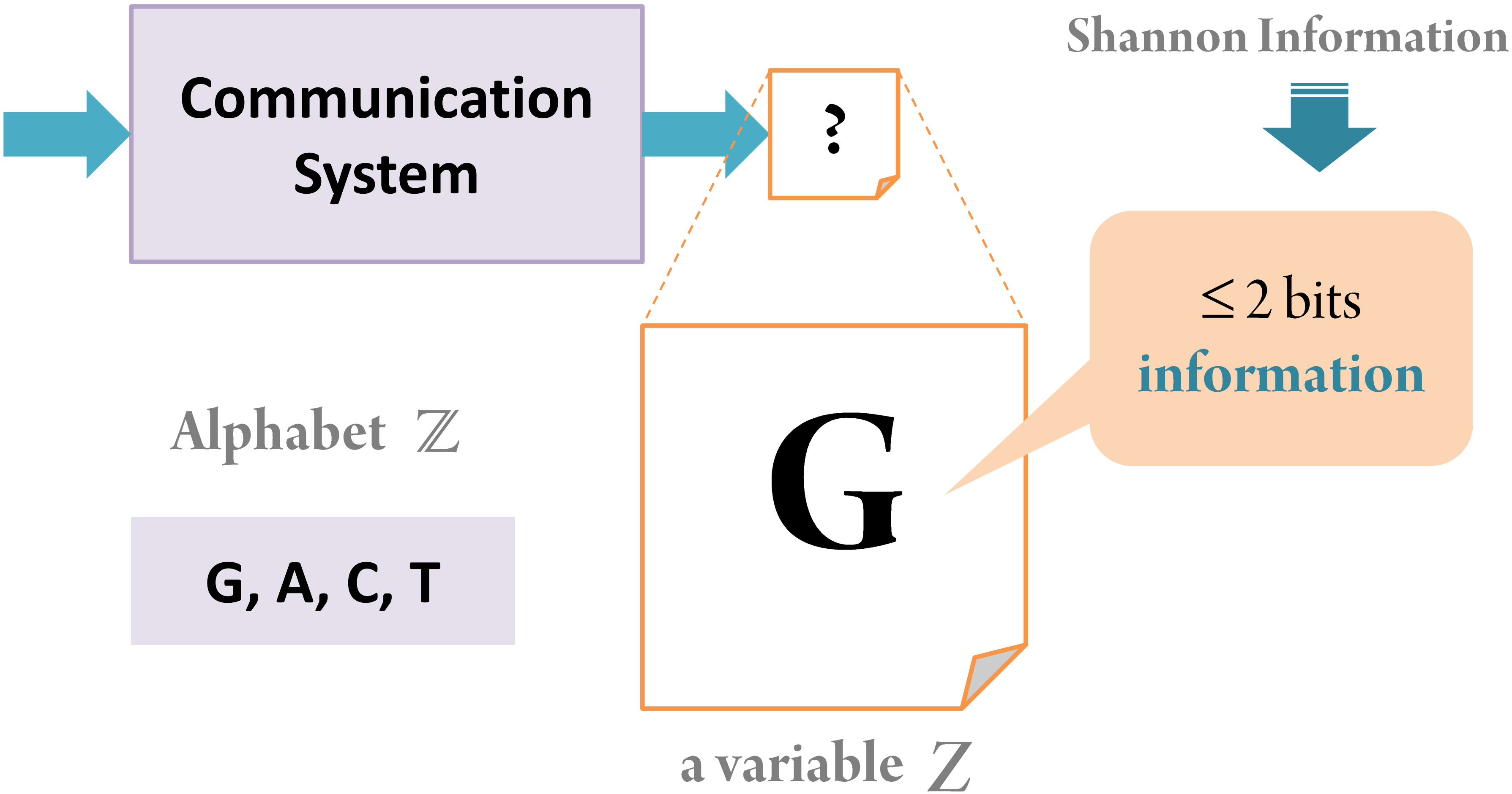
Min Chen, **Cost-Benefit Analysis of Data Intelligence,**
https://vimeo.com/145258513, 2015

$$\mathcal{H}(Z) = -\sum_{z \in Z} \mathrm{p}(z) \log_2 \mathrm{p}(z)$$



Shannon's Definition of Information

$$\mathcal{H}(Z) = -\sum_{z \in Z} \mathrm{p}(z) \log_2 \mathrm{p}(z)$$



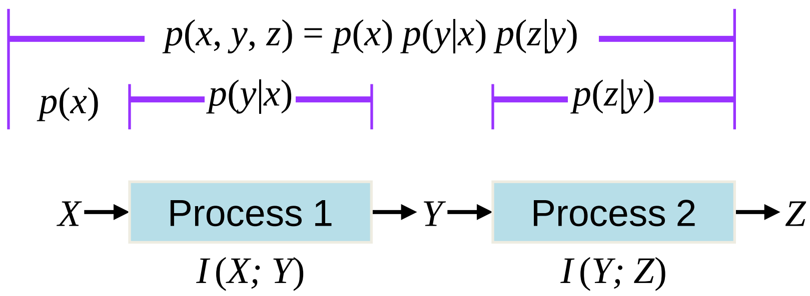
An Alphabet and its Letters

- English alphabet — A, B, C, ..., X, Y, Z
- All English prefixes — bio, geo, pre, pro, ..., un
- All English words — a, ..., silicosis, ..., titin, ...
- All sentences in a text corpus
- All published BioVis papers — ...
- ...

45 letters
pneumonoultramicroscopicsilicovolcanoconiosis

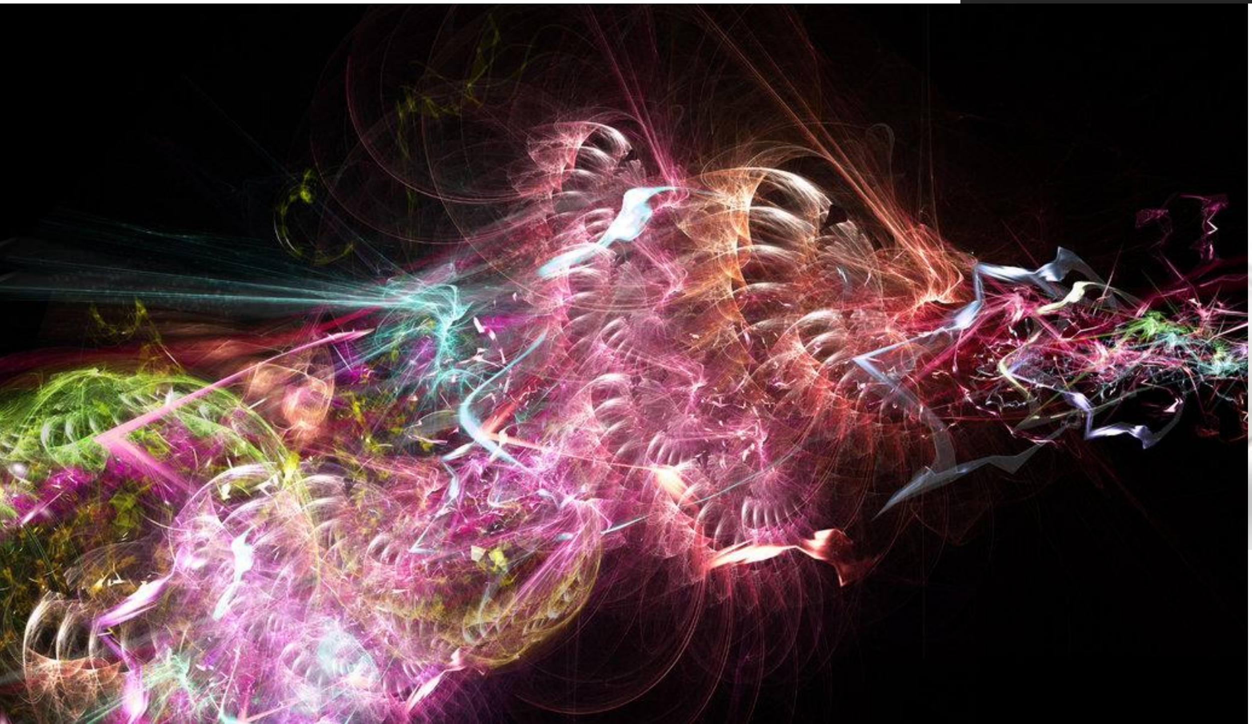
189,819 letters
a word or a formula?

Data Processing Inequality

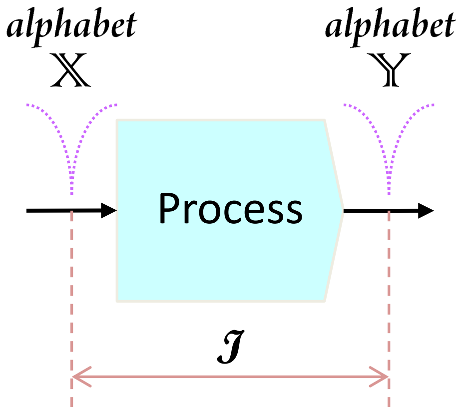


$$I(X;Y) \geq I(X;Z)$$

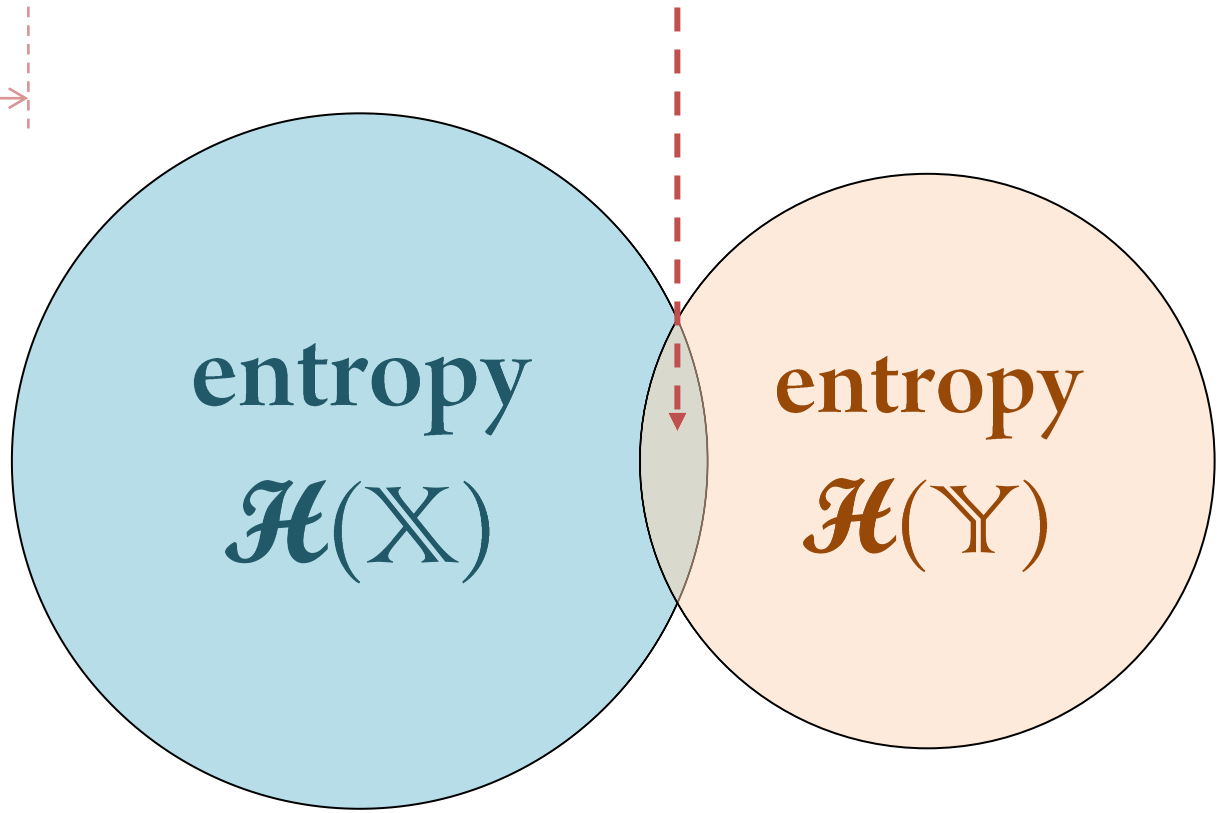
- "No clever manipulation of data can improve the inferences that can be made from the data" [Cover and Thomas, 2006]



Mutual Information (shared uncertainty)



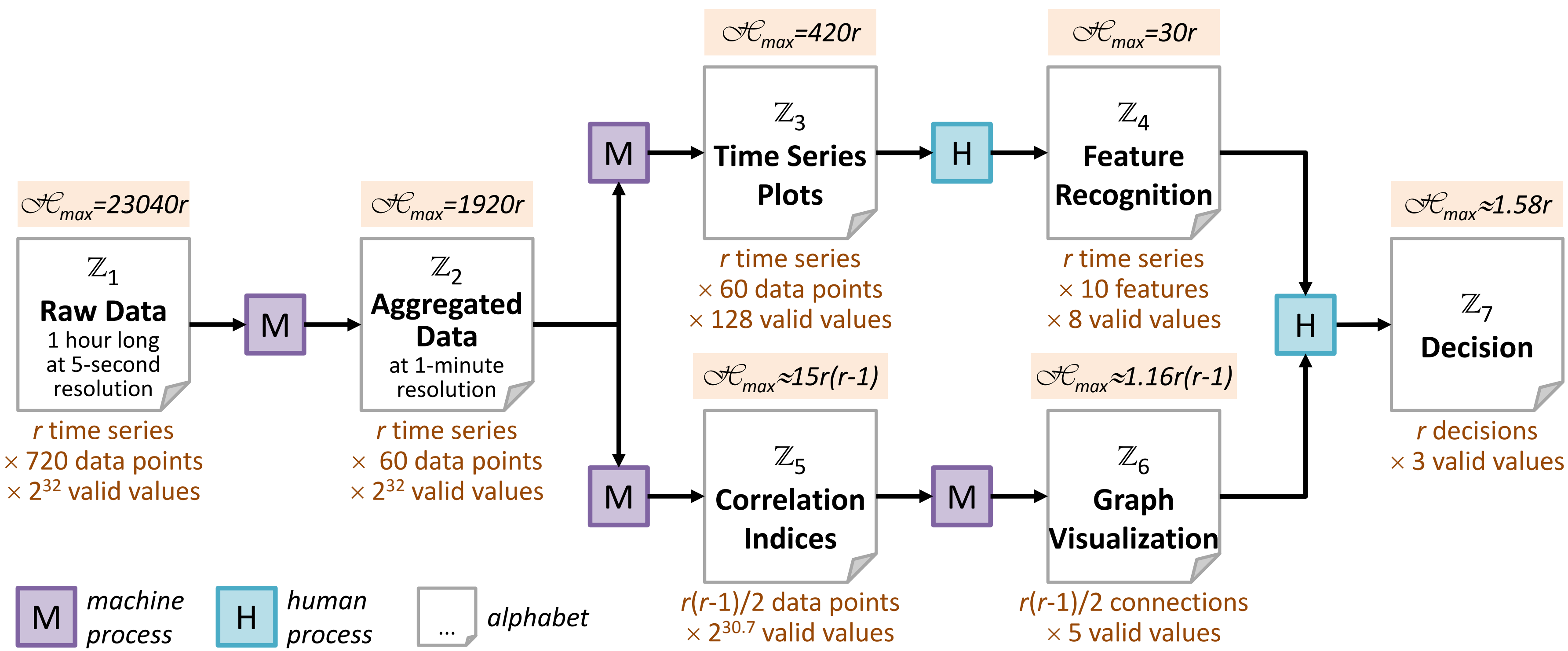
$$\mathcal{I}(X;Y) = \sum_{x\in X}\sum_{y\in Y} \mathbf{p}(x,y)\log_2 \frac{\mathbf{p}(x,y)}{\mathbf{p}_{\mathbf{X}}(x)\mathbf{p}_{\mathbf{Y}}(y)}$$



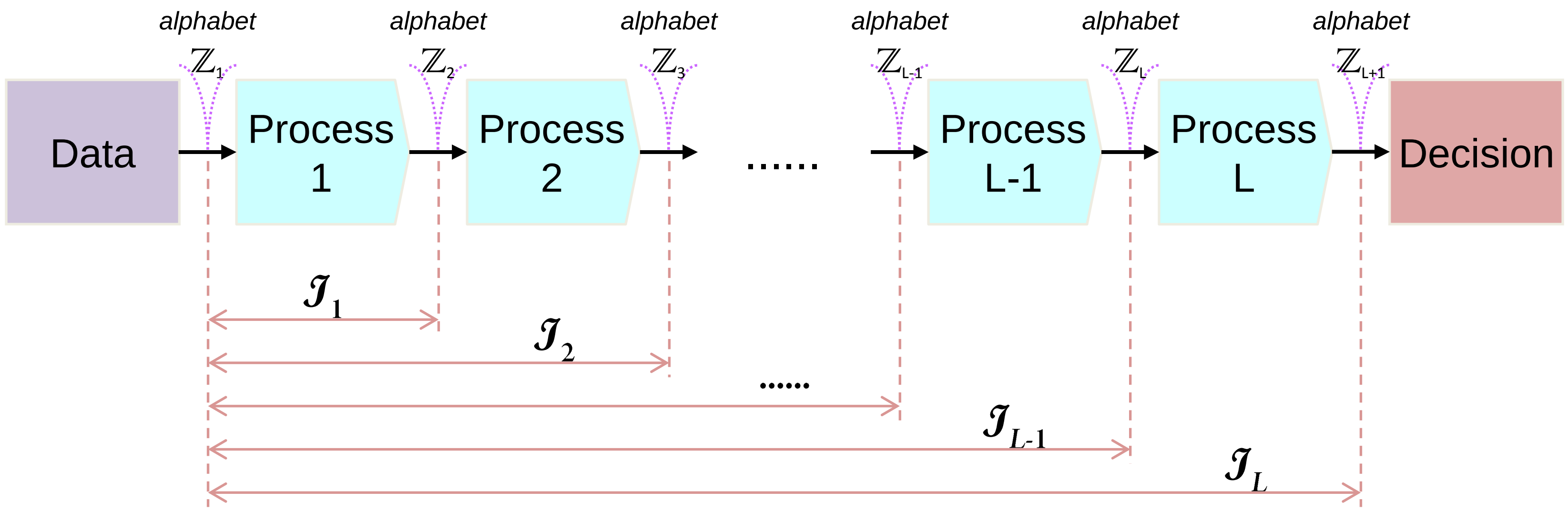
An Example Data Analysis and Visualization Process

- r time series
- 720 data point each series
- 2^{32} valid value each point

- r decisions
- 3 valid values each (e.g., buy, sell, hold)



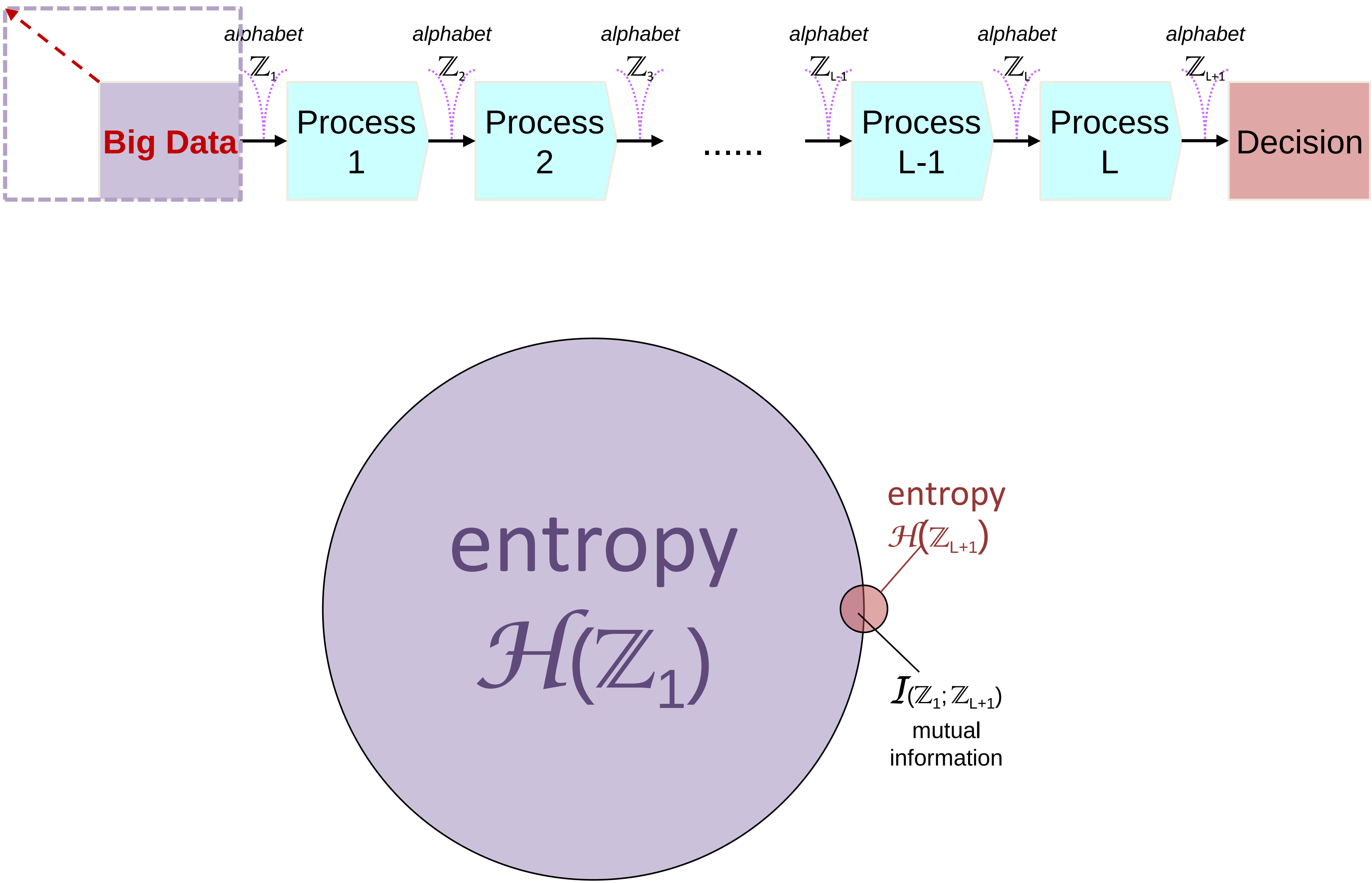
Data Processing Inequality



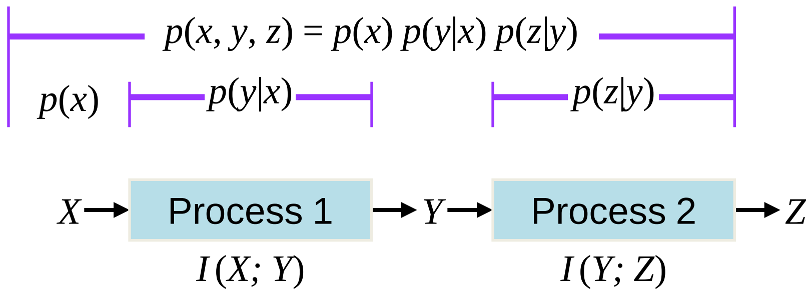
Decreasing of Mutual Information

$$\mathcal{I}_1 \geq \mathcal{I}_2 \geq \ldots \geq \mathcal{I}_{L-1} \geq \mathcal{I}_L$$

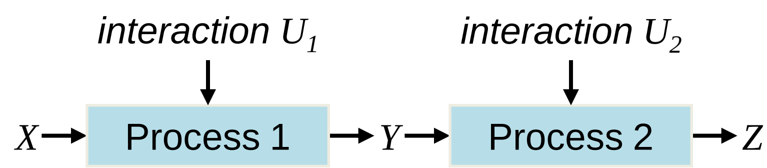
Data Processing Inequality: Big Data Input?

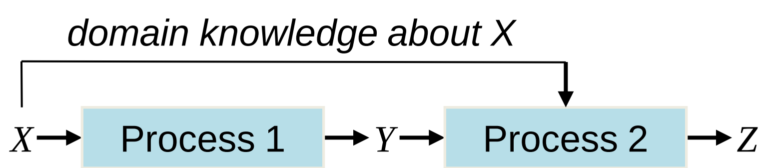


DPI is not Ubiquitous



$$I(X;Y) \geq I(X;Z)$$



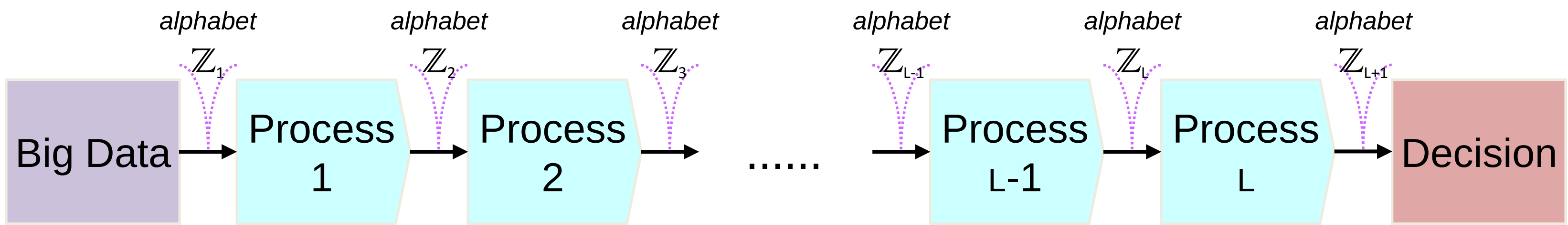


$$I(X;Y) \not\geq I(X;Z)$$

- Markov chain conditions
  - Closed coupling: (X, Y), (Y,Z)
  - X and Z are conditionally independent
- What if one of the conditions is broken?
- In visual analytics, both conditions are usually broken.

M. Chen and H. Jänicke, **An information-theoretic framework for visualization**, *IEEE Transactions on Visualisation and Computer Graphics*, 2010

Soft Knowledge in Data Intelligence



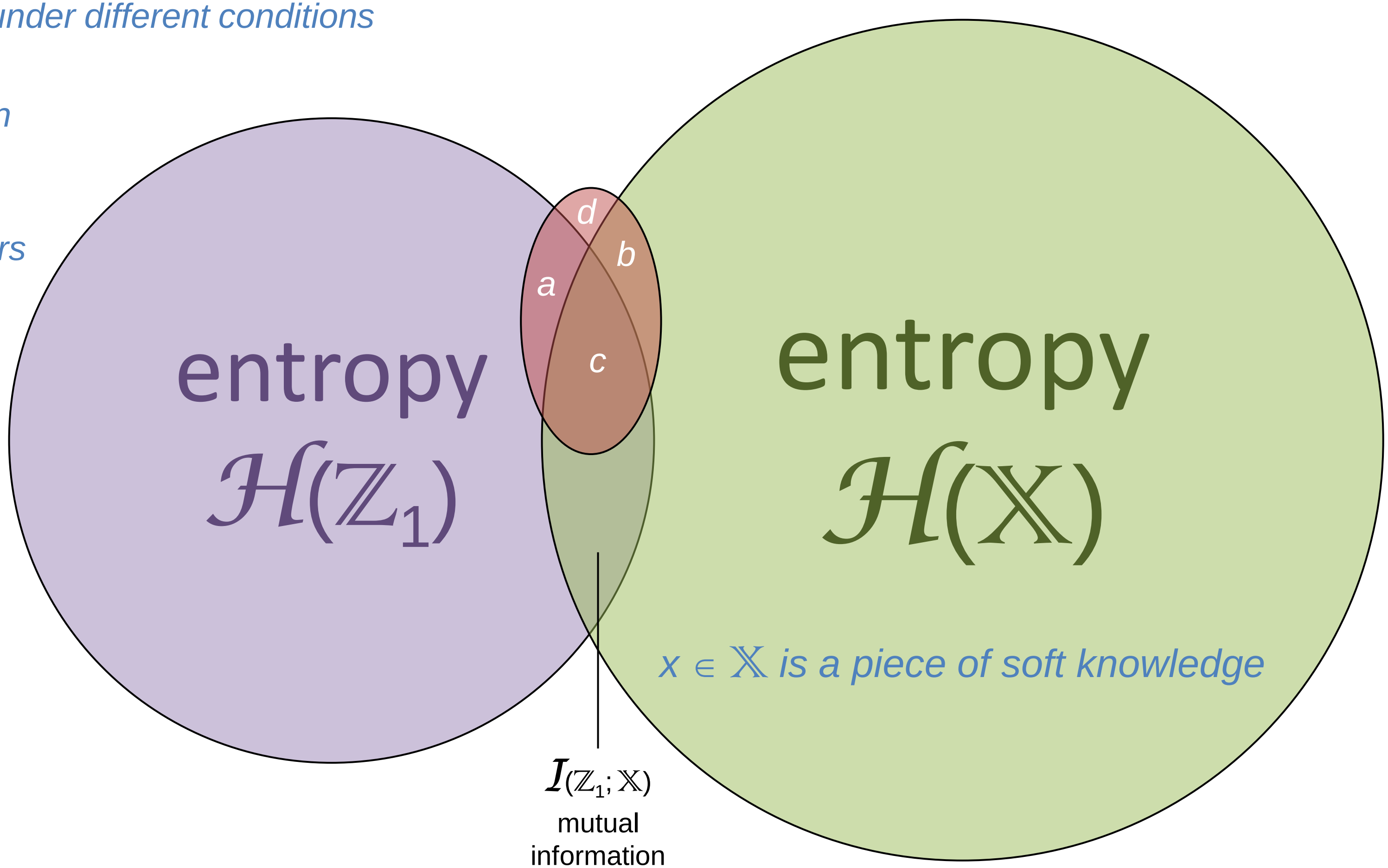
All possible decisions under different conditions

a) totally data-driven

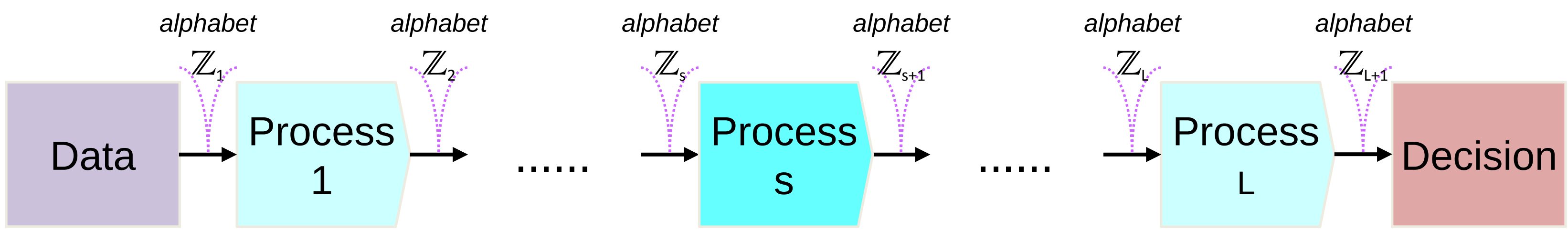
b) totally instinct-driven

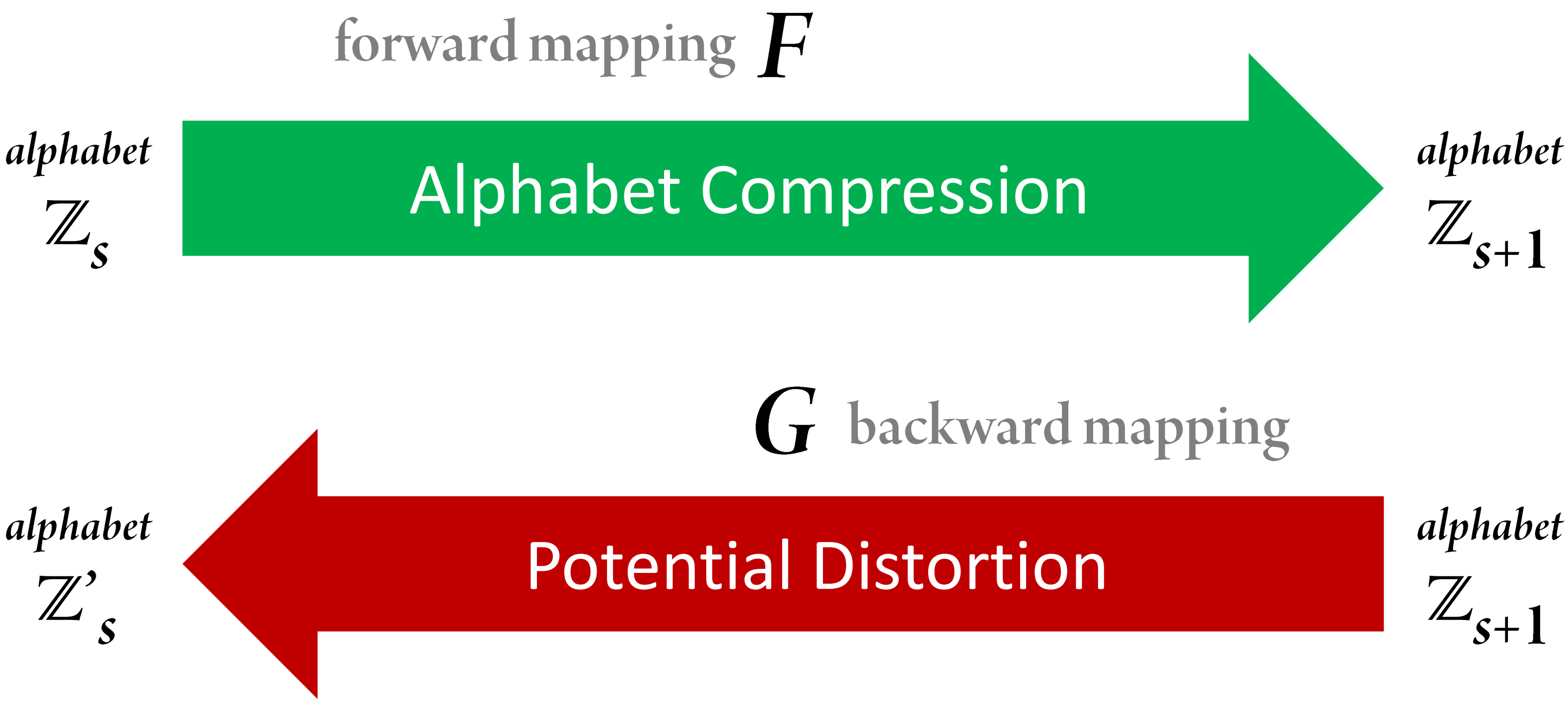
c) data-informed

d) due to unknown or uncontrollable factors



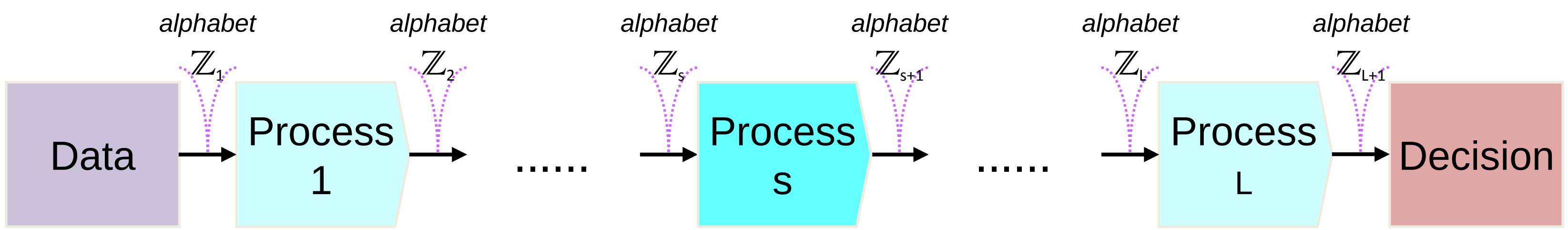
Transformation





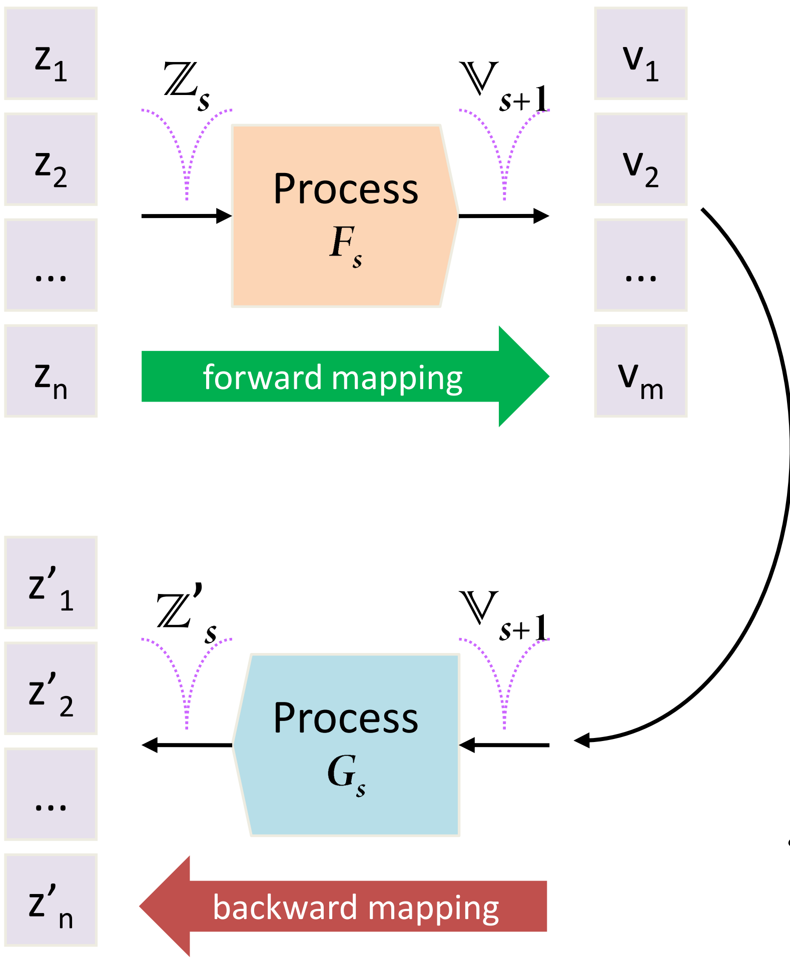
M. Chen and A. Golan, **What may visualization processes optimize?**, *IEEE Transactions on Visualisation and Computer Graphics*, 2015

A Sequential Workflow and Two Basic Metrics



- The s^{th} Function (Process): $F_s : \mathbb{Z}_s \longrightarrow \mathbb{Z}_{s+1}$
- Alphabetic Compression Ratio (ACR): $\Psi_{ACR}(F_s) = \frac{\mathscr{H}(Z_{s+1})}{\mathscr{H}(Z_s)}$
- A Reverse "Guessing" Process: $G_s : \mathbb{Z}_{s+1} \longrightarrow \mathbb{Z}'_s$
- Potential Distortion Ratio (PDR): $\Psi_{PDR}(F_s) = \frac{\mathscr{D}_{KL}(Z'_s||Z_s)}{\mathscr{H}(Z_s)}$

Kullback–Leibler divergence

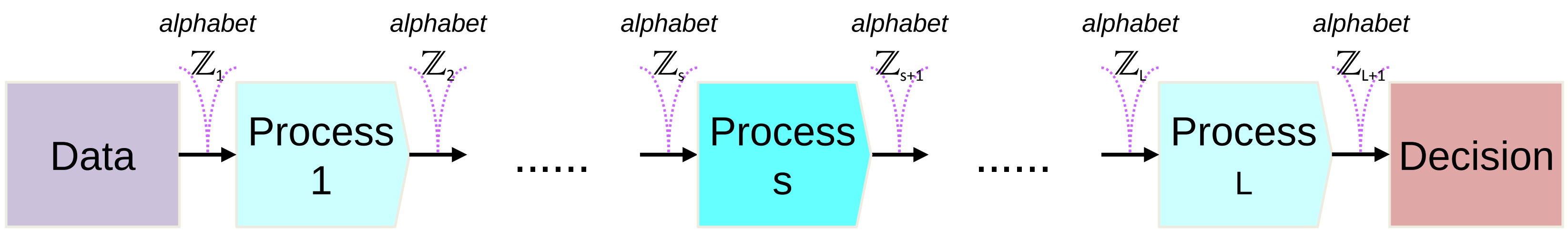


Solomon Kullback
1907-1994

Richard Leibler
1914-2003

$$\mathcal{D}_{KL}(Z'||Z) = \sum_{(z=z')\in Z} \mathrm{p}(z') \log_2 \frac{\mathbf{p}(z')}{\mathbf{q}(z)}$$

Cost-Benefit Ratio



- Effectual Compression Ratio (ECR):

$$\Psi_{ECR}(F_s) = \frac{\mathscr{H}(\mathbb{Z}_{s+1}) + \mathscr{D}_{KL}(\mathbb{Z}'_s || \mathbb{Z}_s)}{\mathscr{H}(\mathbb{Z}_s)}$$

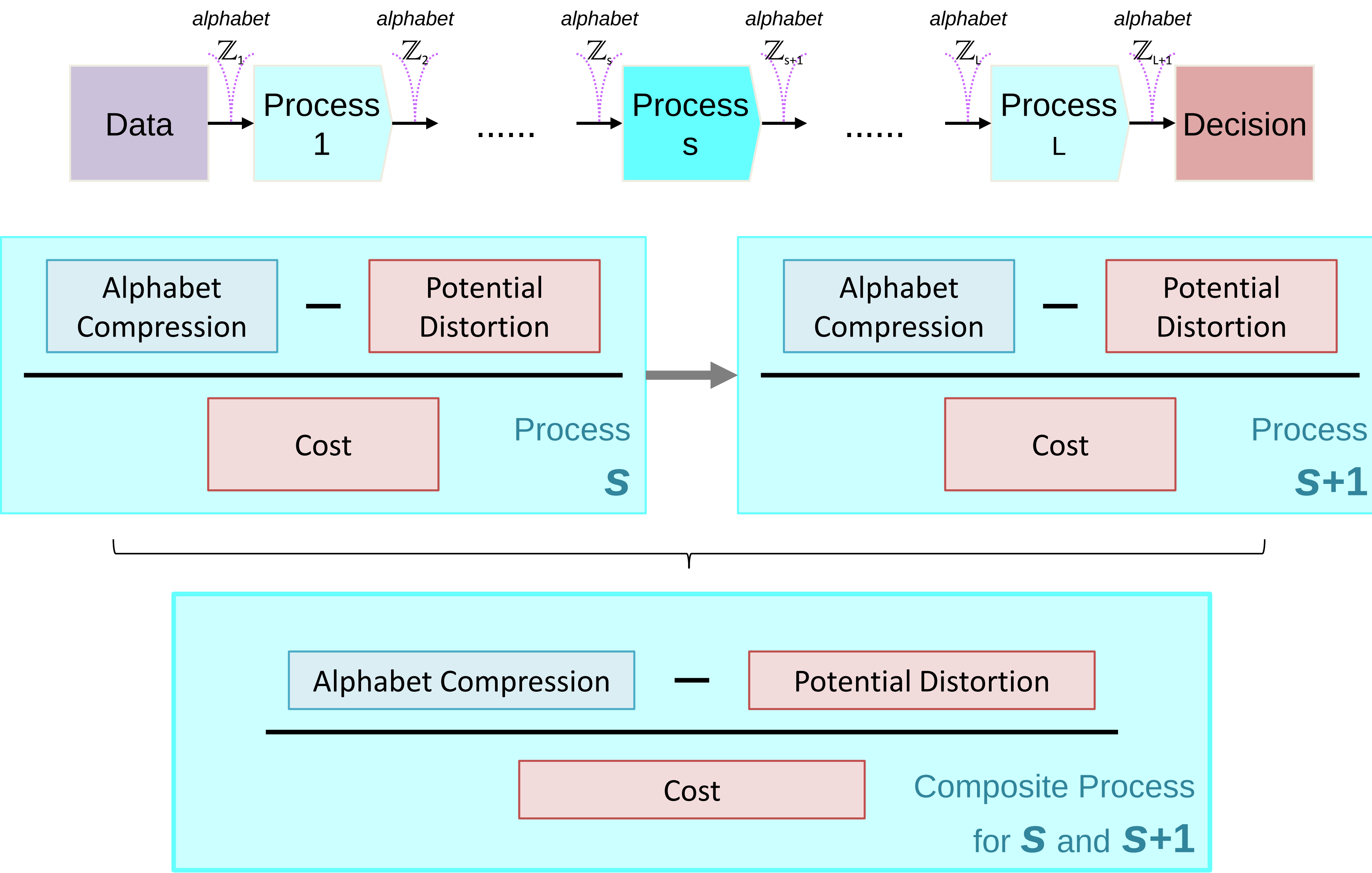
- Incremental Cost-Benefit Ratio (ICBR):

$$\Upsilon(F_s) = \frac{\mathscr{B}(F_s)}{\mathscr{C}(F_s)} = \frac{\mathscr{H}(\mathbb{Z}_s) - \mathscr{H}(\mathbb{Z}_{s+1}) - \mathscr{D}_{KL}(\mathbb{Z}'_s || \mathbb{Z}_s)}{\mathscr{C}(F_s)}$$

- *Cost can be measured in energy, time, money, etc.*

M. Chen and A. Golan, **What may visualization processes optimize?**, *IEEE Transactions on Visualization and Computer Graphics*, 2015

Cost-Benefit Optimization



Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

Three Spaces and Three Measures

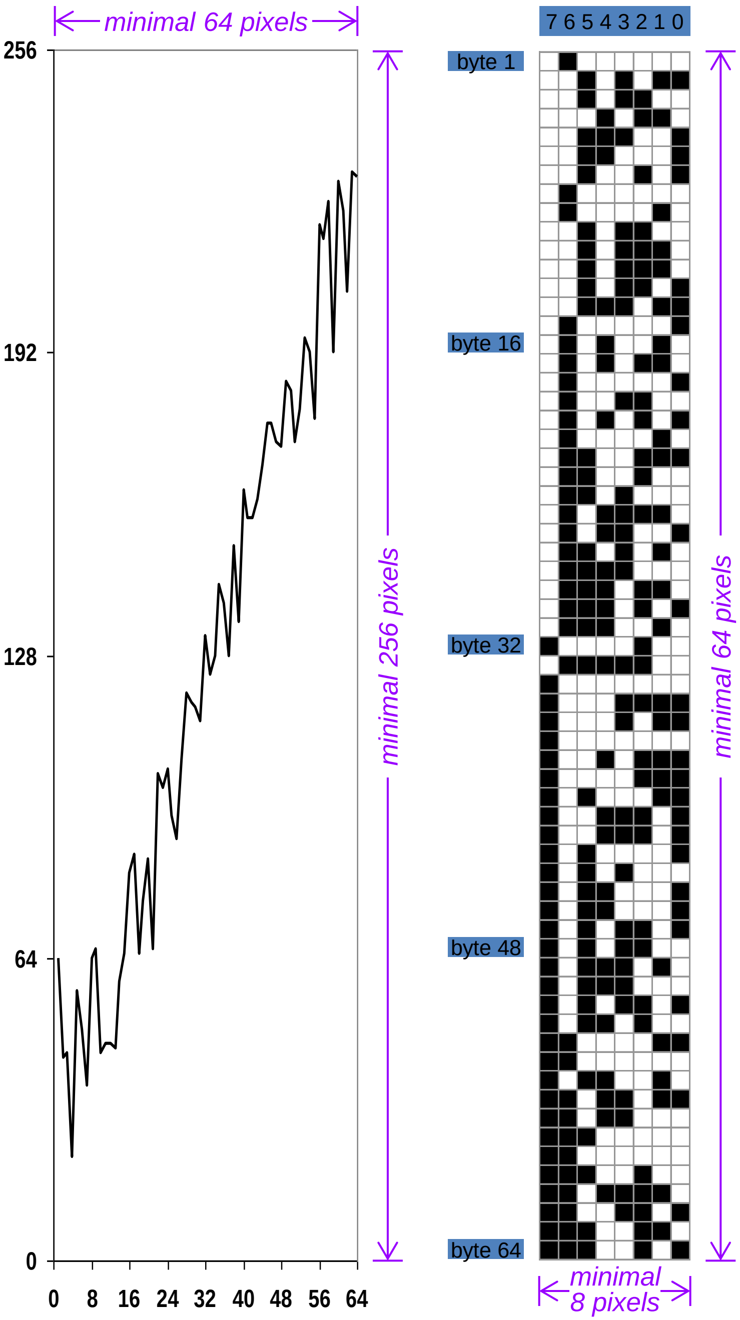
M. Chen and H. Jänicke, **An information-theoretic framework for visualization**, *IEEE Transactions on Visualisation and Computer Graphics*, 2010

- Entropy of Input Data Space: $\boldsymbol{H}(X)$
- Visualization Capacity: $\boldsymbol{V}(G)$
- Display Capacity: $\boldsymbol{D}$

$$\text{Visual Mapping Ratio (VMR)} = \frac{\boldsymbol{V}(G)}{\boldsymbol{H}(X)}$$

$$\text{Information Loss Ratio (ILR)} = \frac{\max(\boldsymbol{H}(X) - \boldsymbol{V}(G), 0)}{\boldsymbol{H}(X)}$$

$$\text{Display Space Utilization (DSU)} = \frac{\boldsymbol{V}(G)}{\boldsymbol{D}}$$



Example of V(G)

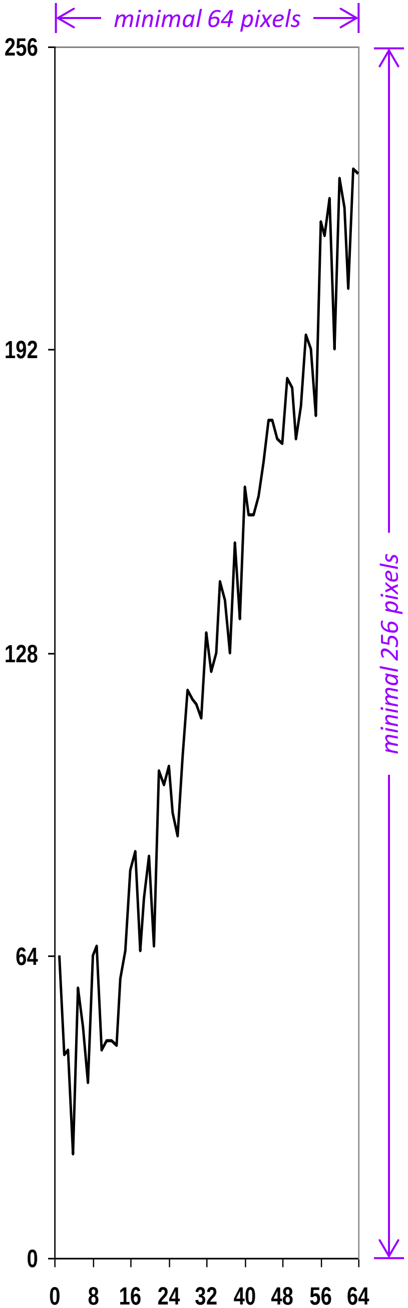
- Entropy of Data Alphabet

$$\mathbf{H}(Z) = -\sum_{t=0}^{64}\sum_{i=0}^{255}\frac{1}{256}\log_2\frac{1}{256} = 512$$

- $\boldsymbol{V}(G) = \boldsymbol{H}(Z)$
- Binary Pixel Plot: $\boldsymbol{D}$
  - 4x4 pixels per bit
  - D: $2^9 \times 2^4 = 2^{13}$ bits
- Time Series Plot: $\boldsymbol{D}$
  - Minimal 256x64 pixels
  - D: $2^8 \times 2^6 = 2^{14}$ bits
- The more compact, the better?
- Cost?
- Reconstructability?

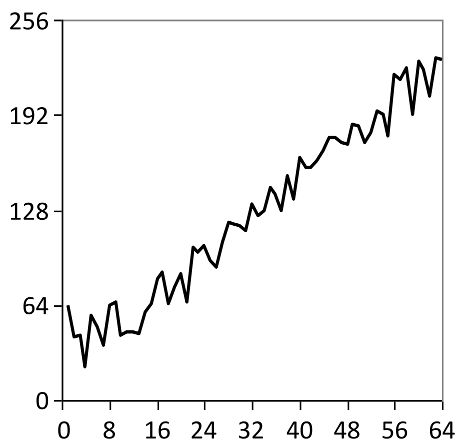
Information Loss Ratio (ILR)

- Display Space Restriction
  - 64x64 pixels
- Evenly distributed probability mass function
- Linear visual mapping
- ILR is a probabilistic measure about
  - a data space *X*
  - not an instance x_i



H(X) = 512 bits
V(G) = 512 bits
D = 16382 bits
VMR = 1
ILR = 0
DSU = 0.03

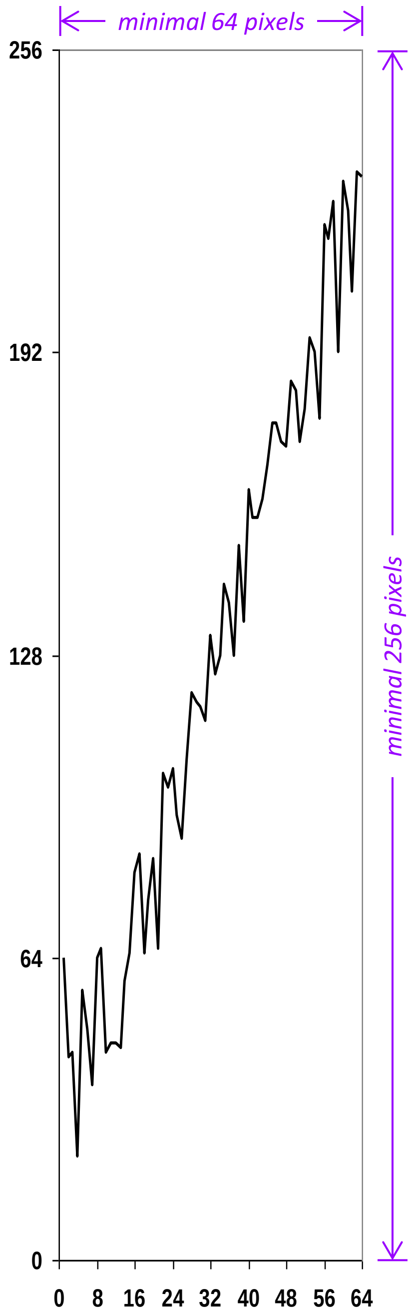
H(X) = 512 bits
V(G) = 384 bits
D = 4096 bits
VMR = 0.75
ILR = 0.25
DSU = 0.09



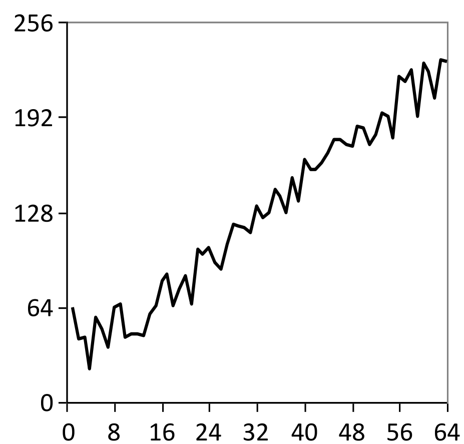
(a) evenly distributed *p*

Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary



information loss:
25%



(a) evenly distributed p

Information Loss Ratio (ILR)

- Display Space Restriction
  - 64x64 pixels
- Evenly distributed probability mass function
- Linear visual mapping
- ILR is a probabilistic measure about
  - a data space X
  - not an instance x_i

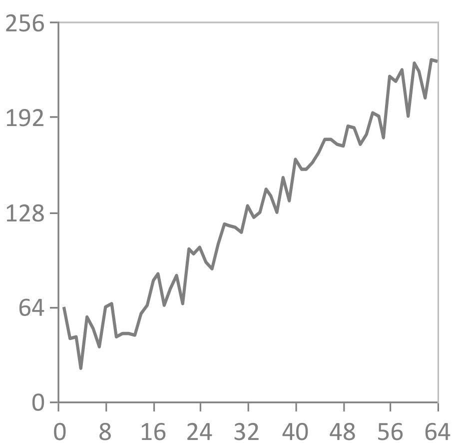
M. Chen and H. Jänicke, **An information-theoretic framework for visualization**, *IEEE Transactions on Visualisation and Computer Graphics*, 2010

Non-uniform Distribution

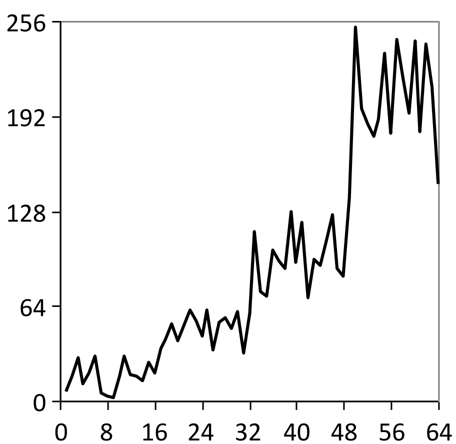
- Linear visual mapping

information loss:
25.0%

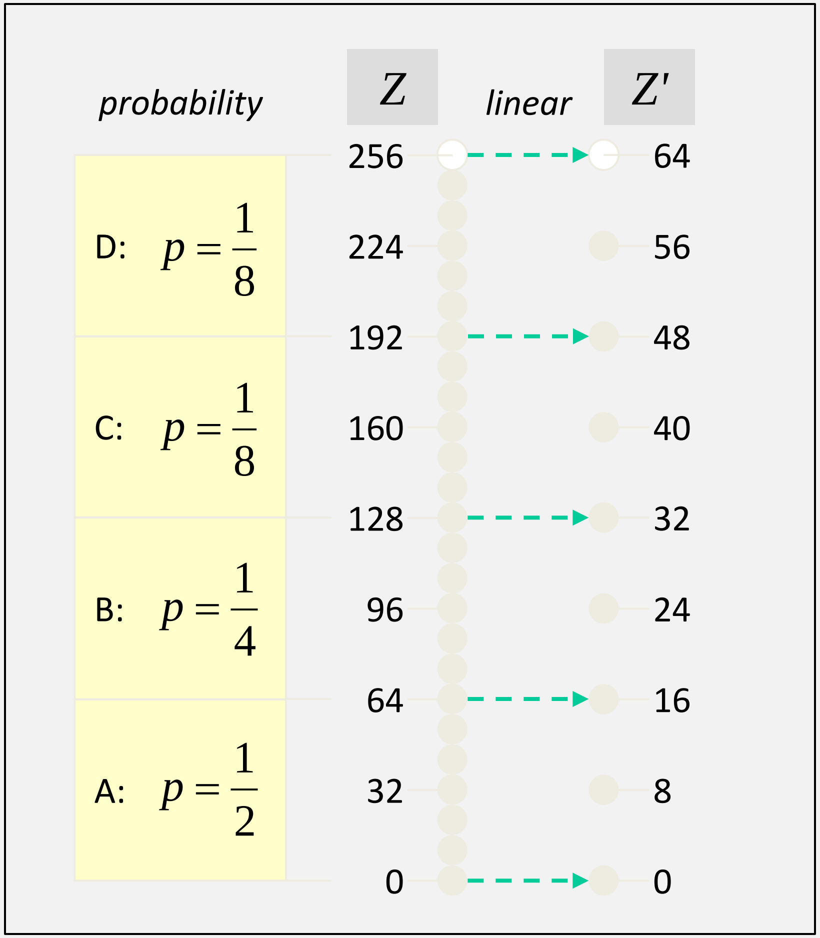
information loss:
25.8%



(a) evenly distributed *p*



(b) unevenly distributed *p*



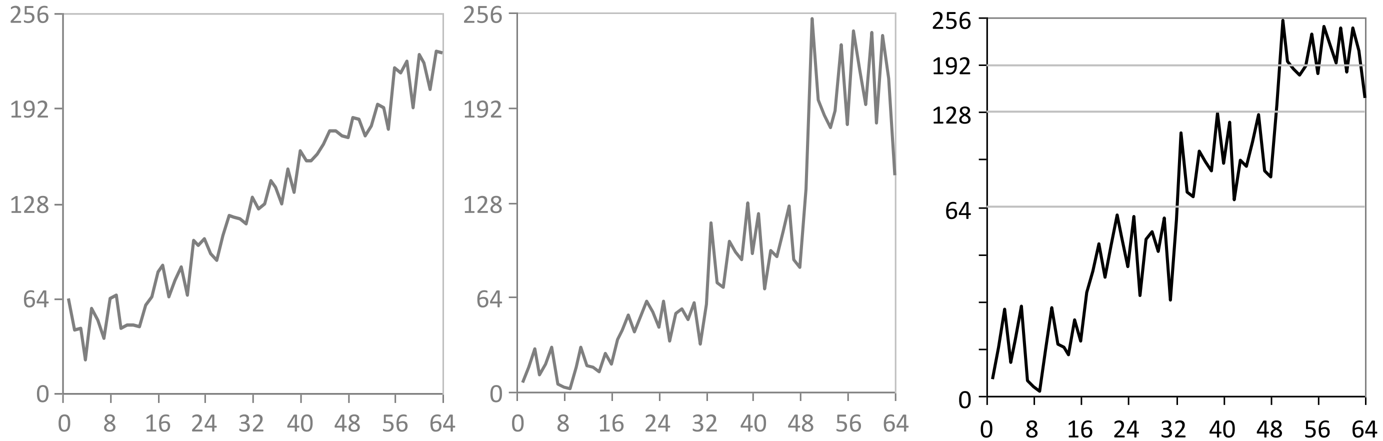
Non-uniform Distribution

- Nonlinear visual mapping

information loss:
25.0%

information loss:
25.8%

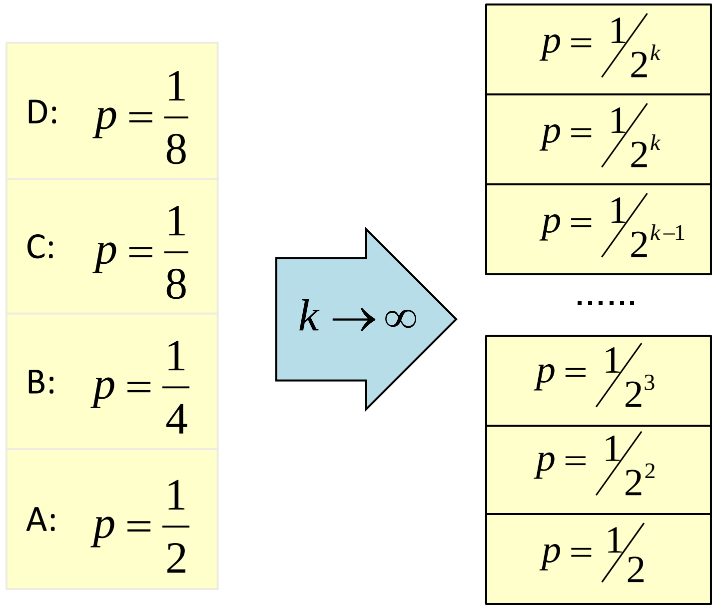
information loss:
22.6%



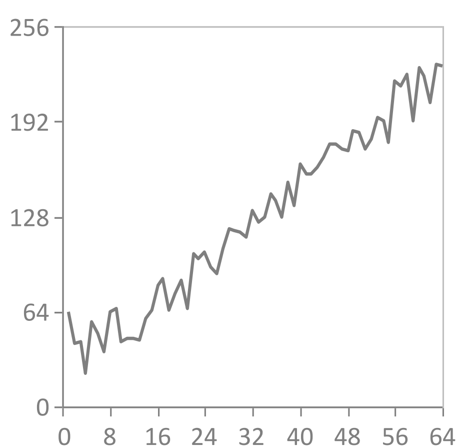
(a) evenly distributed *p* (b) unevenly distributed *p* (c) 4 regional mappings

Non-uniform Distribution

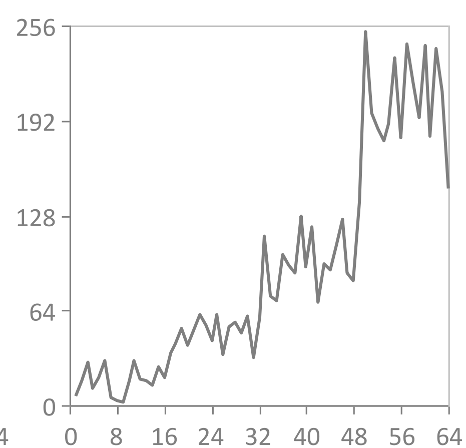
- Logarithmic visual mapping



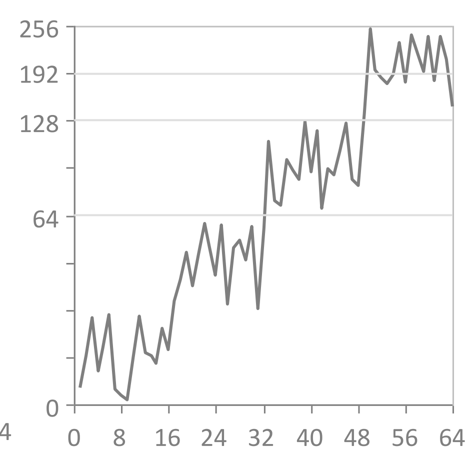
information loss:	information loss:	information loss:	information loss:
25.0%	25.8%	22.6%	0%



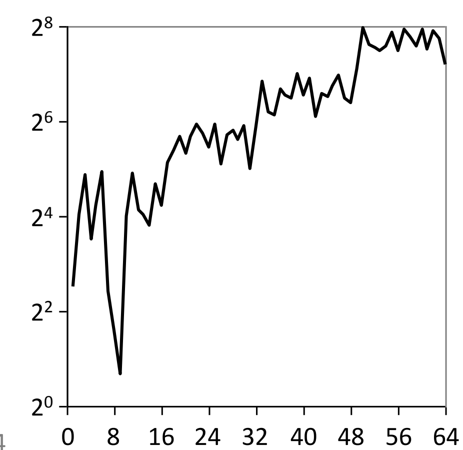
(a) evenly distributed *p*



(b) unevenly distributed *p*



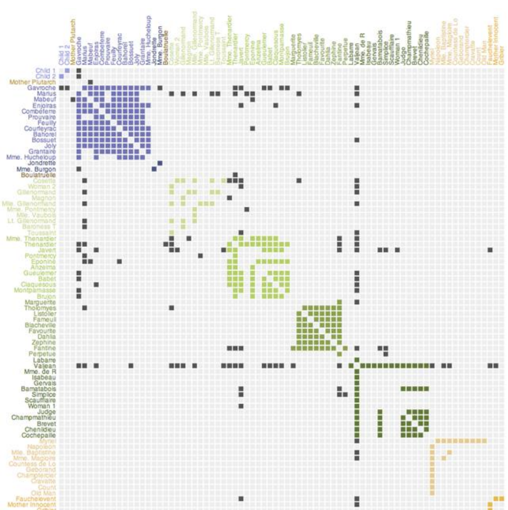
(c) 4 regional mappings

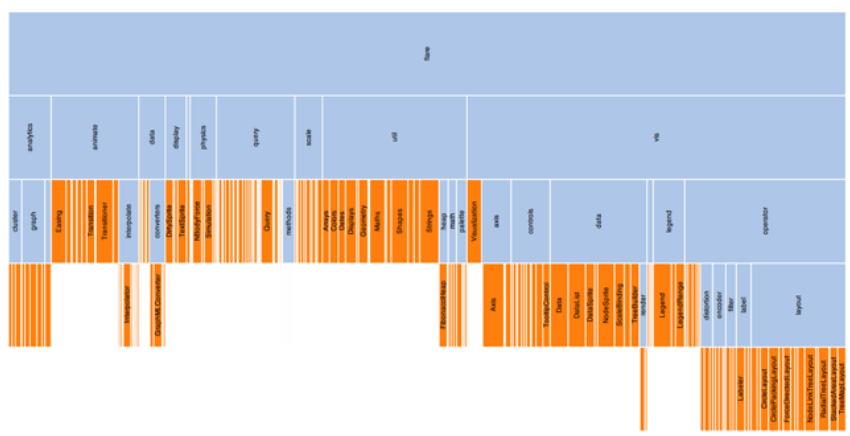


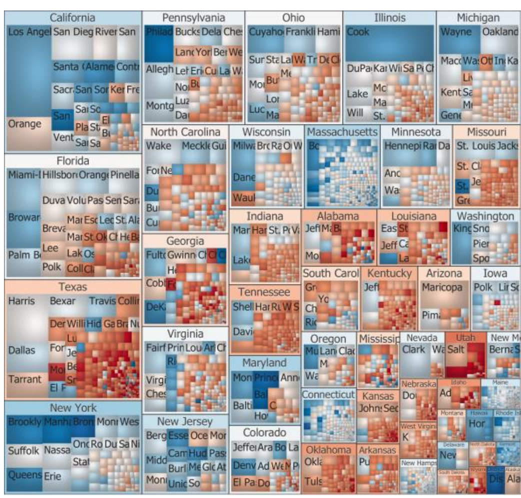
(d) logarithmic plot

A Prediction?

- Display capacity D is limited.
- Data space: noticeable non-uniform distribution: $H(X) << H_{max}(X)$
- Visualization capacity $V(G)$: a visual representation exhibiting a similar non-uniform distribution of space requirement can reduce information loss.
- Can adjacency matrixces be improved for trees?







http://hci.stanford.edu/jheer/files/zoo/

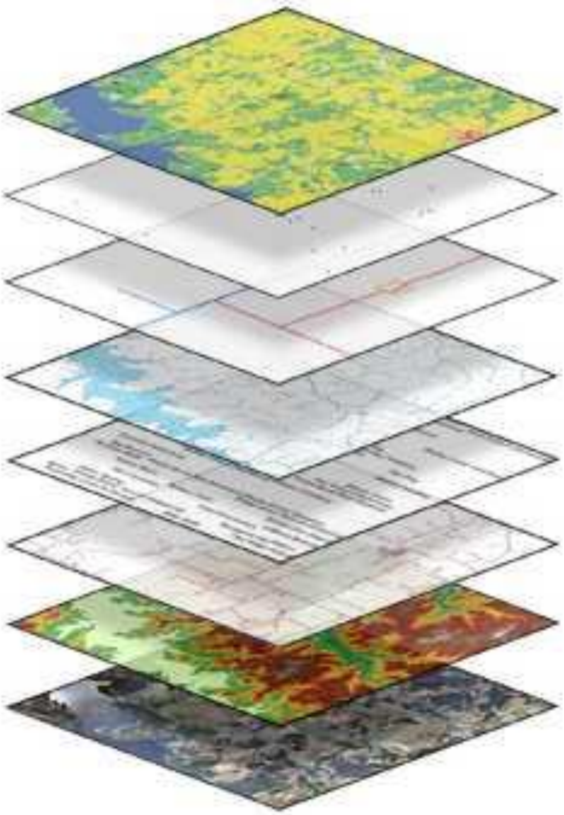
http://en.wikipedia.org/wiki/Treemapping

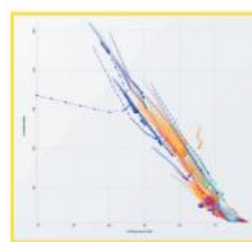
Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

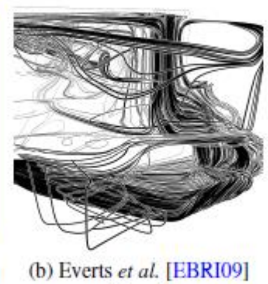
Map Overlay and its Generalization in Visualization

http://learnpracticalgis.com/how-to-overlay-maps/

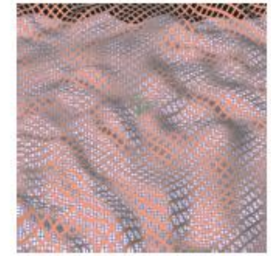




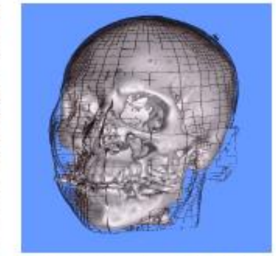
(a) Robertson *et al.* [RFF*08]
Type B



(b) Everts *et al.* [EBRI09]
Type C

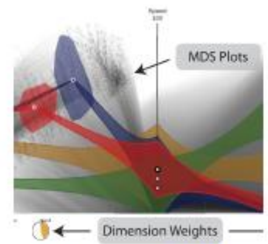


(c) Bair, House [BH07]
Types C, D, G

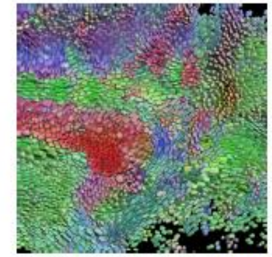


(d) Treavett, Chen [TC00]
Type D, J

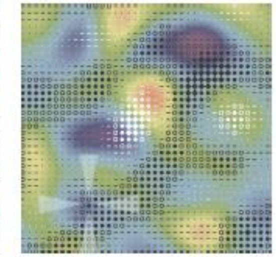
(e) Collins *et al.* [CPC09]
Types C, E



(f) Guo *et al.* [GXY12]
Type E

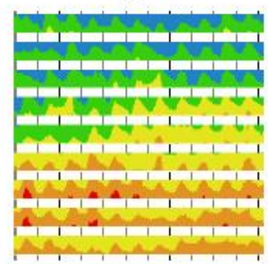


(g) Kindlmann, Westin [KW06]
Type F

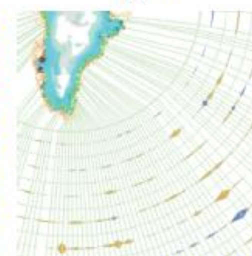


(h) Ware [War09]
Types G, J

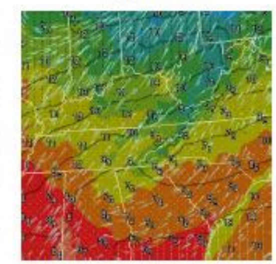
(i) Chen *et al.* [CPL*11]
Types C, G, J



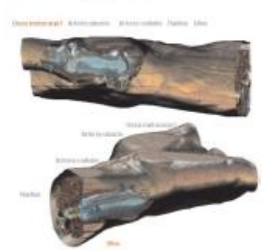
(j) Saito *et al.* [SMY*05]
Type C, H



(k) Drocourt *et al.* [DBS*11]
Type H

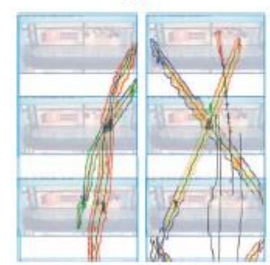


(l) Ware, Plumlee [WP13]
Type I

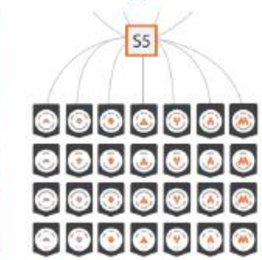


(m) Viola *et al.* [VFSG06]
Type E, J

(n) Correa *et al.* [CSC06]
Type H, J



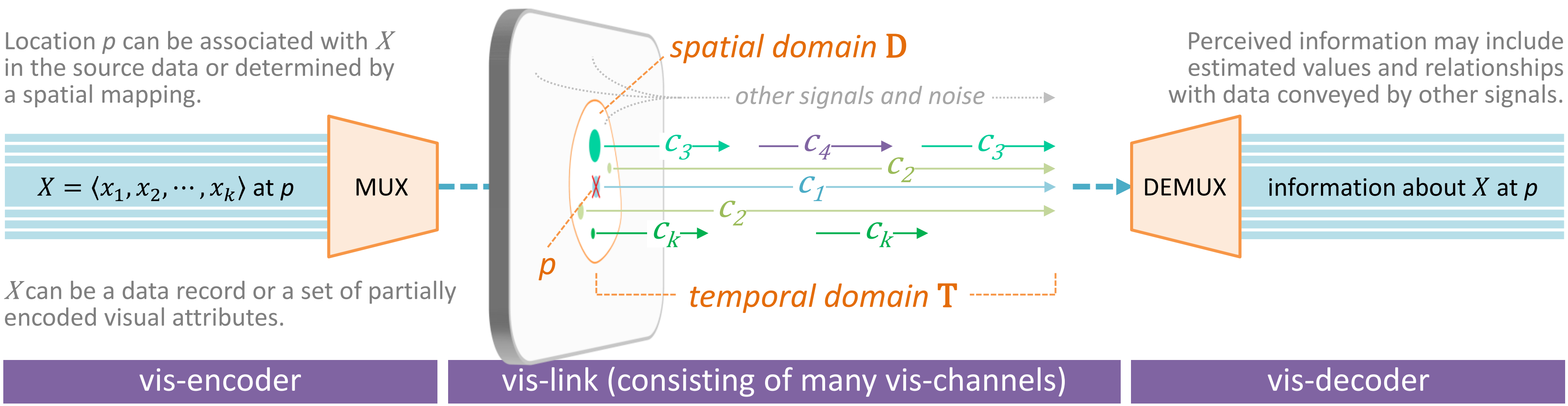
(o) Botchen *et al.* [BBS*08]
Types E, G, H, J



(p) Maguire *et al.* [MRSS*12]
Type J

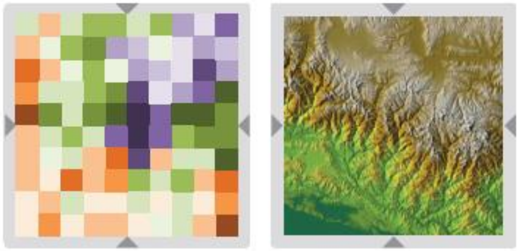
Visual Multiplexing

- Frequency-division multiplexing (FDM)
- Time-division multiplexing (TDM)
- Space-division multiplexing (SDM)
- Code-division multiplexing (CDM)

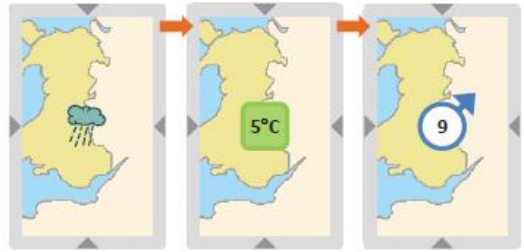


M. Chen, S. Walton, K. Berger, J. Thiyagalingam, B. Duffy, H. Fang, C. Holloway, and A. E. Trefethen, **Visual multiplexing**, *Computer Graphics Forum*, 2014.

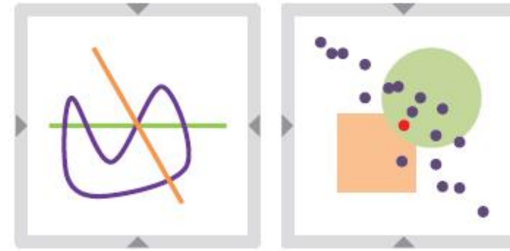
10 Types of Visual Multiplexing



(a) Type A: Partition a space

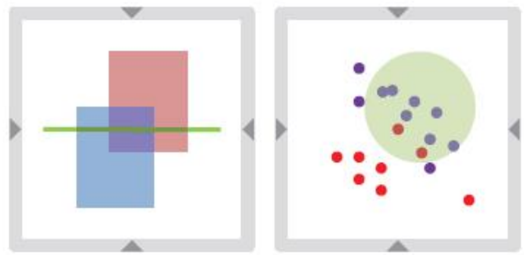


(b) Type B: Partition a time period

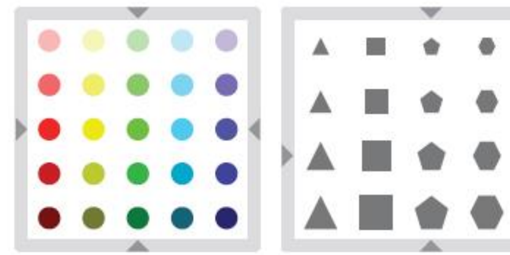


(c) Type C: Introduce partial occlusion

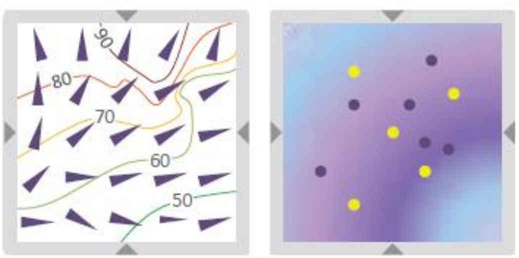
(d) Type D: Use a 'hollow' visual channel



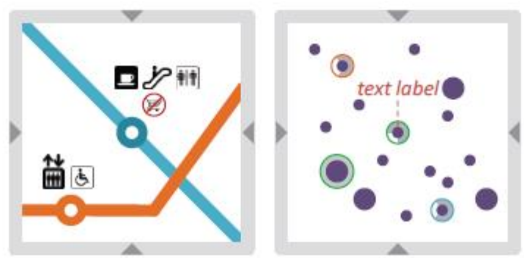
(e) Type E: Introduce translucent occlusion



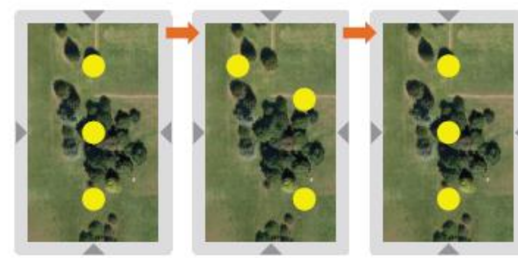
(f) Type F: Use an integrated visual channel



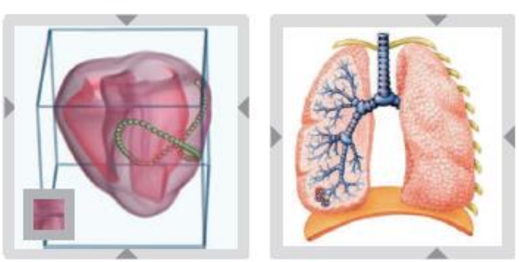
(g) Type G: Depict a continuous field



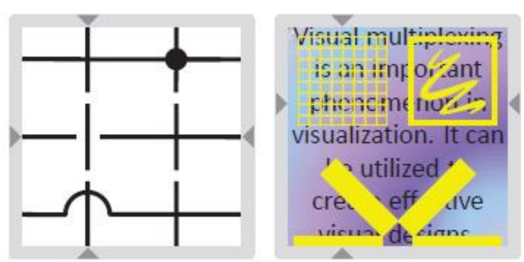
(h) Type H: Shift a visual channel



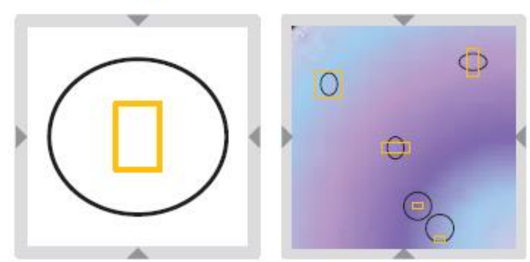
(i) Type I: Use periodic motion



(j) Type J: Assume a priori knowledge

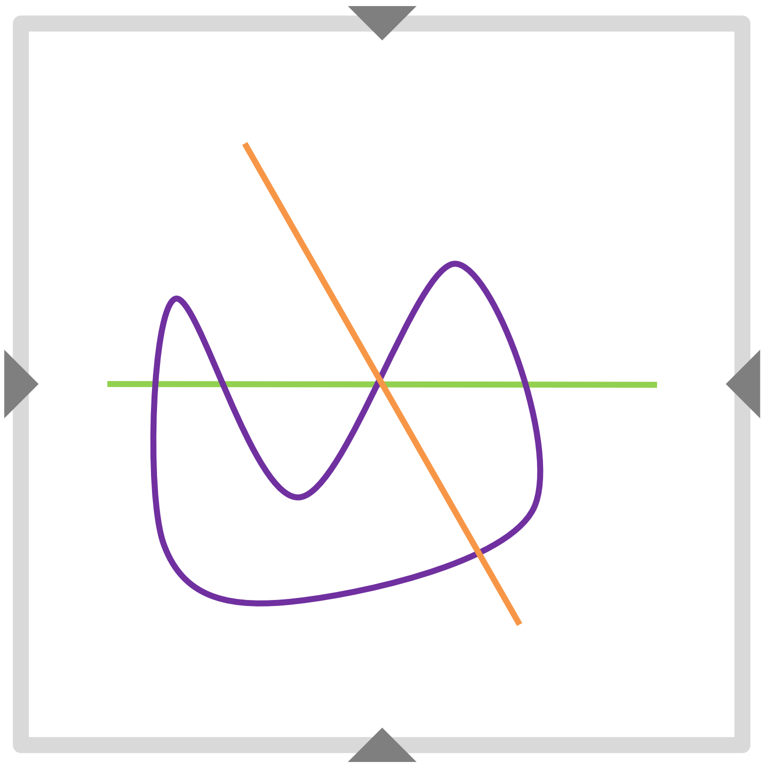


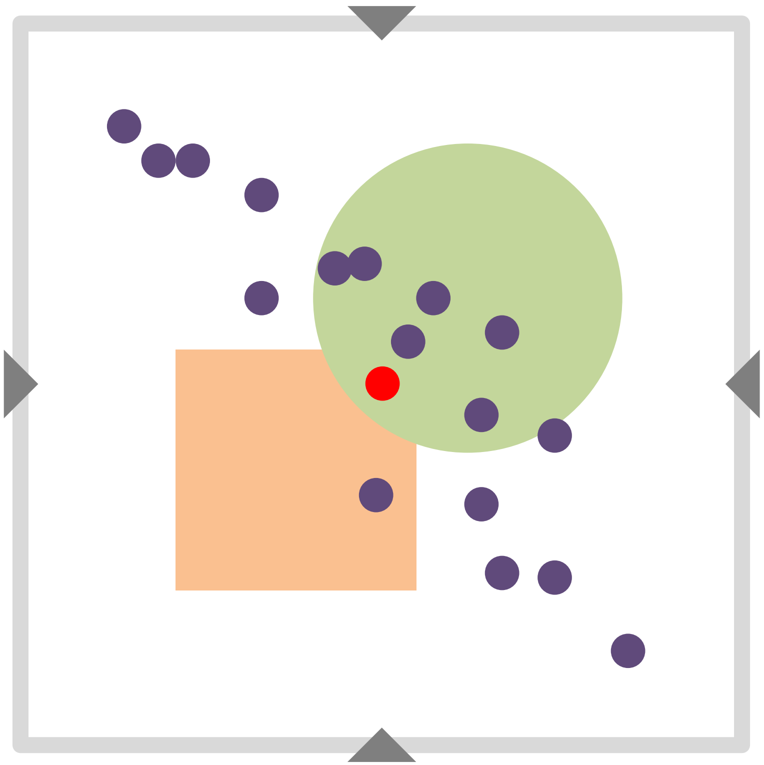
(k) Type J (continued): Acquired knowledge



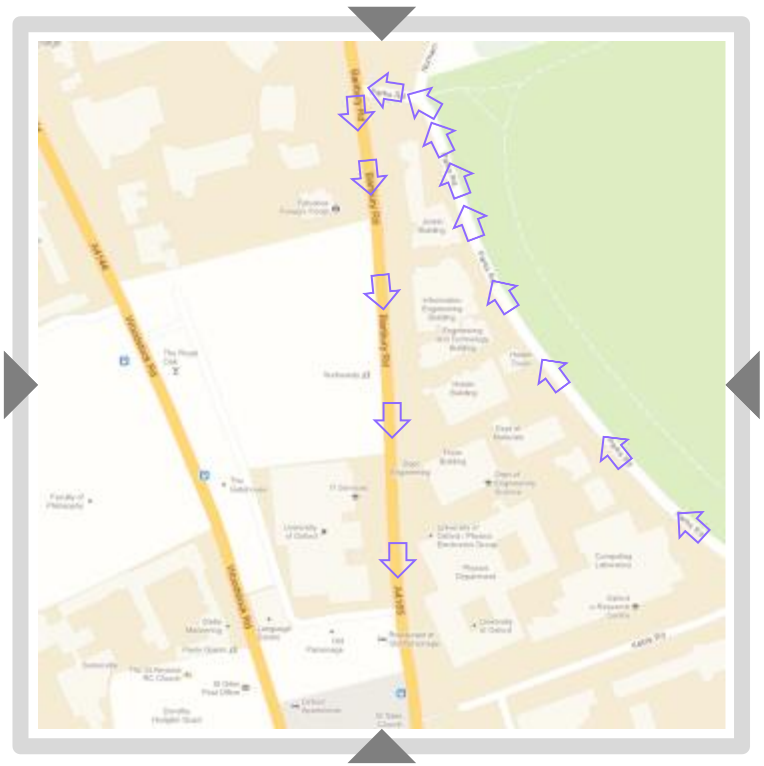
(l) Type J (continued): Visual language

Type C: Introduce Partial Occlusion

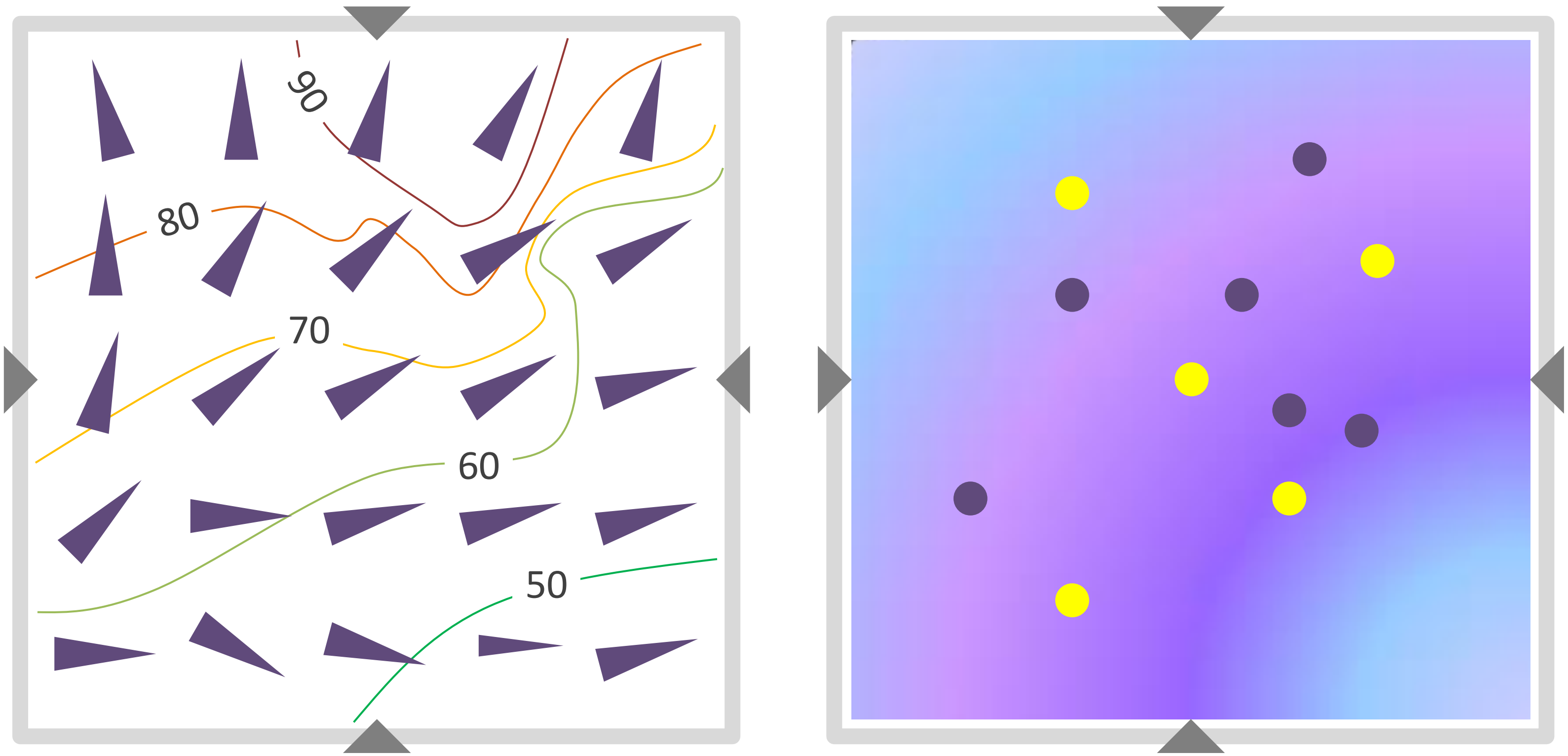




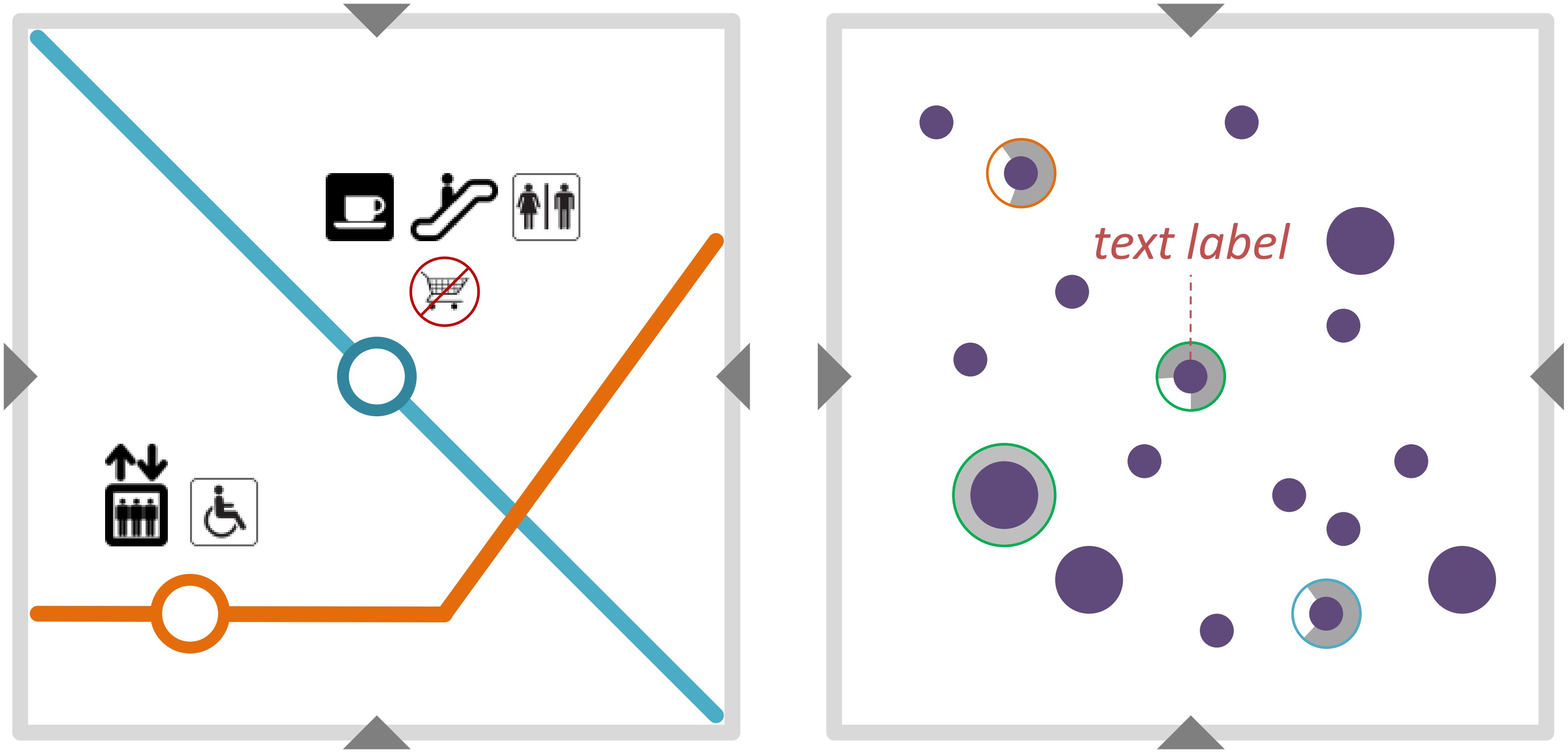
Type D: Use a 'Hollow' Visual Channel



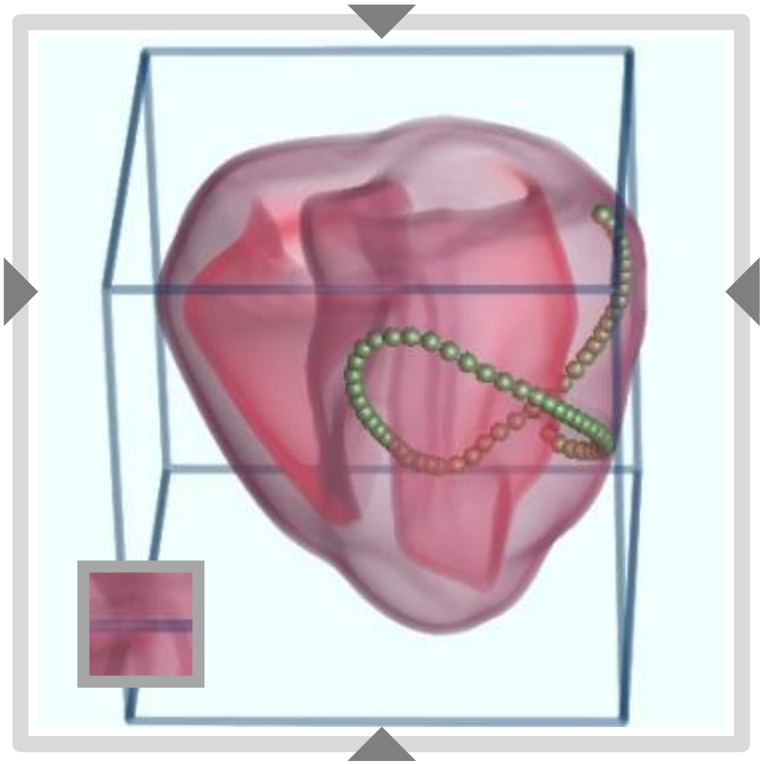
Type G: Depict a Continuous Field

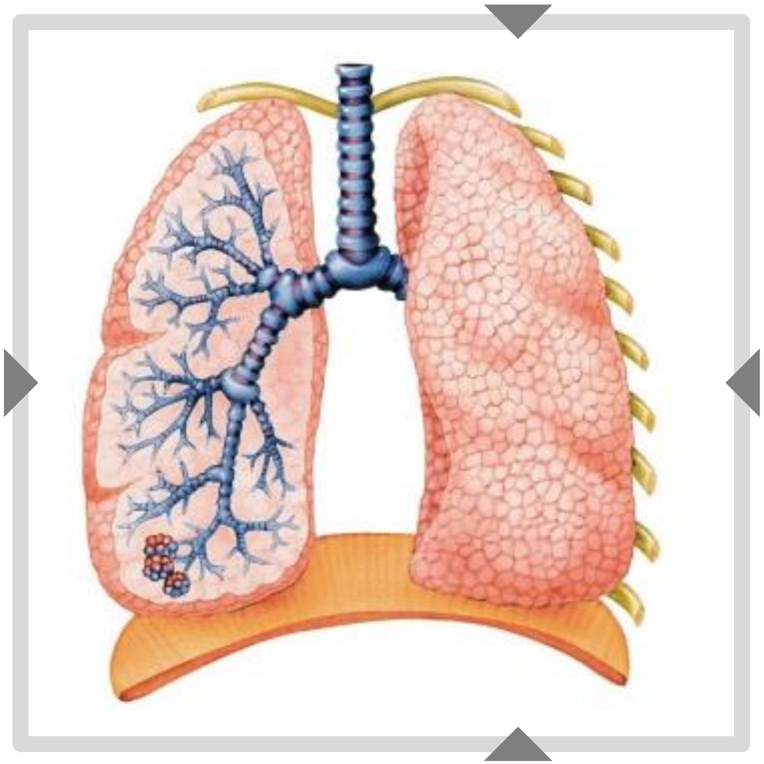


Type H: Shift a Visual Channel

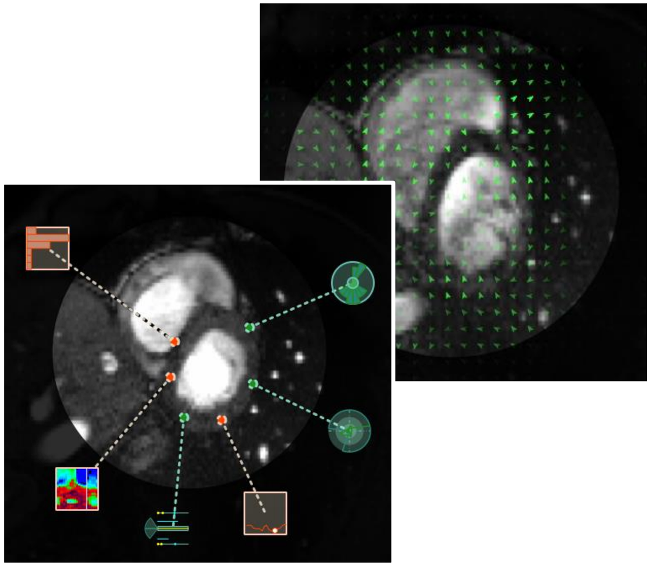


Type J: Assume A Priori Knowledge





How about When Every Pixel is Used?



Data Space → Visualization Space → Display Space

H

Data Space Entropy

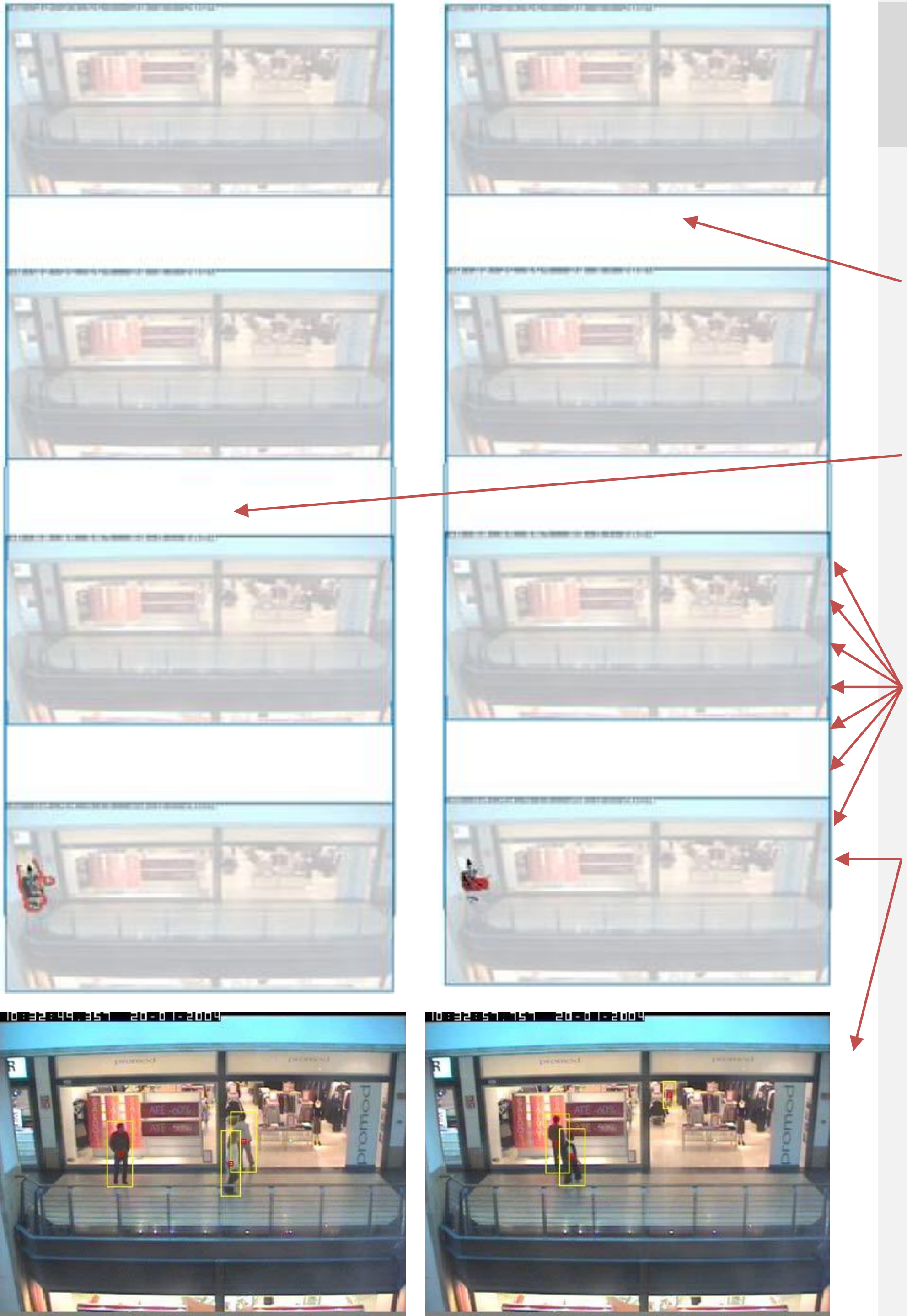
V(G)

Visualization Capacity (Visualization Space Entropy)

D

Display Space Capacity

$$\frac{V(G)}{D} \ll 1$$



Systematics Application

- C: Introduce Partial Occlusion
- E: Introduce Translucent Occlusion
- G: Depict a Continuous Field
- J: Assume A Priori Knowledge

R. P. Botchen, S. Bachthaler, F. Schick, M. Chen, G. Mori, D. Weiskopf and T. Ertl, **Action-based multi-field video visualization**, *IEEE Transactions on Visualization and Computer Graphics*, 2008.

Information Theory and Visualization

1. Data Intelligence —— a big picture
2. Visualization —— a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. **Example: Error Detection and Correction**
6. Example: Process Optimization
7. Summary

Information Theory and Visualization

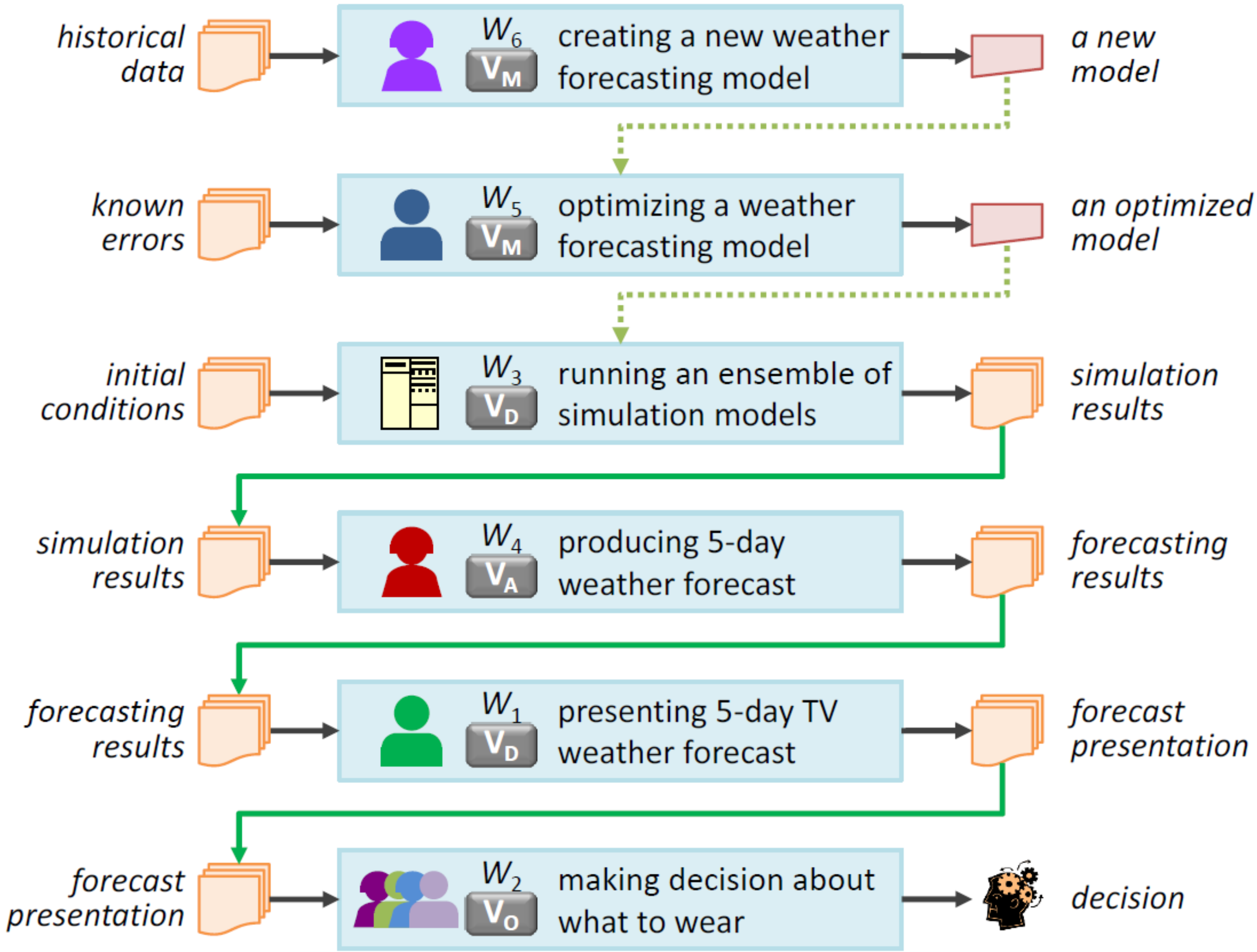
1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

Four Levels of Visualization

1. Disseminative Level — $O(1)$
   - This is “a”!
2. Observational Level — $O(n)$
   - “a”, “b”, “c”, ... what, when, where?
3. Analytical Level — $O(n^k)$
   - Are “a”, “b”, “c” related? Why?
4. Model-developmental Level — $O(kn)$, $O(n!)$
   - How does “a” lead to “b”?

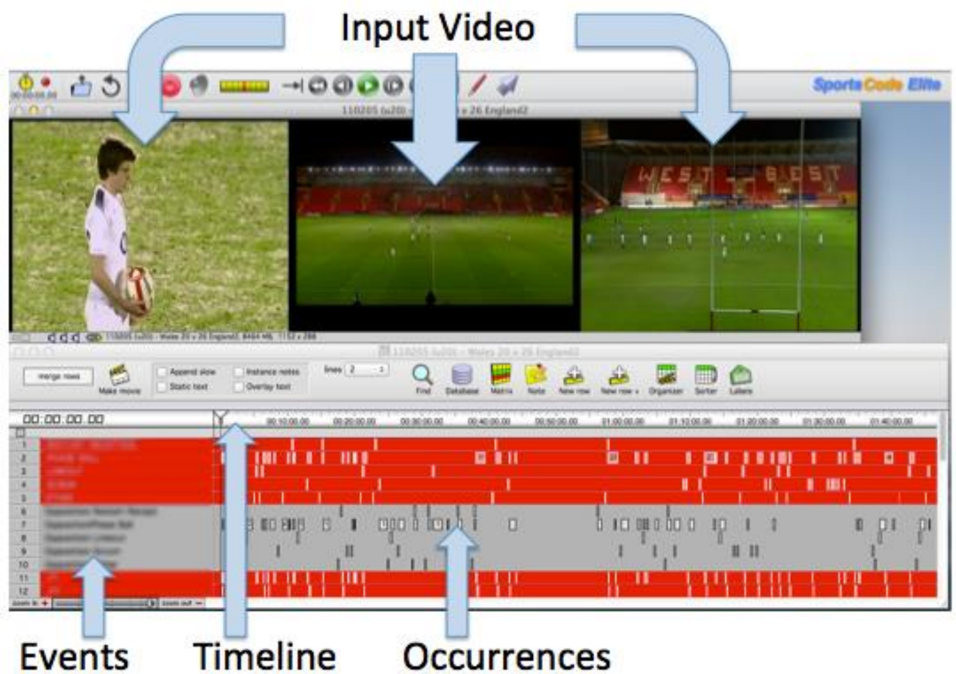
M. Chen and A. Golan, **What may visualization processes optimize?**, *IEEE Transactions on Visualisation and Computer Graphics*, 2015

A Composite Workflow

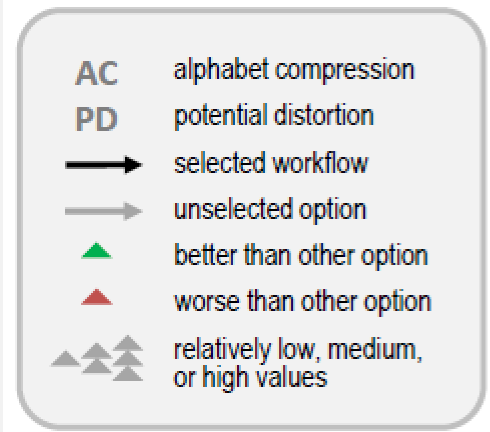


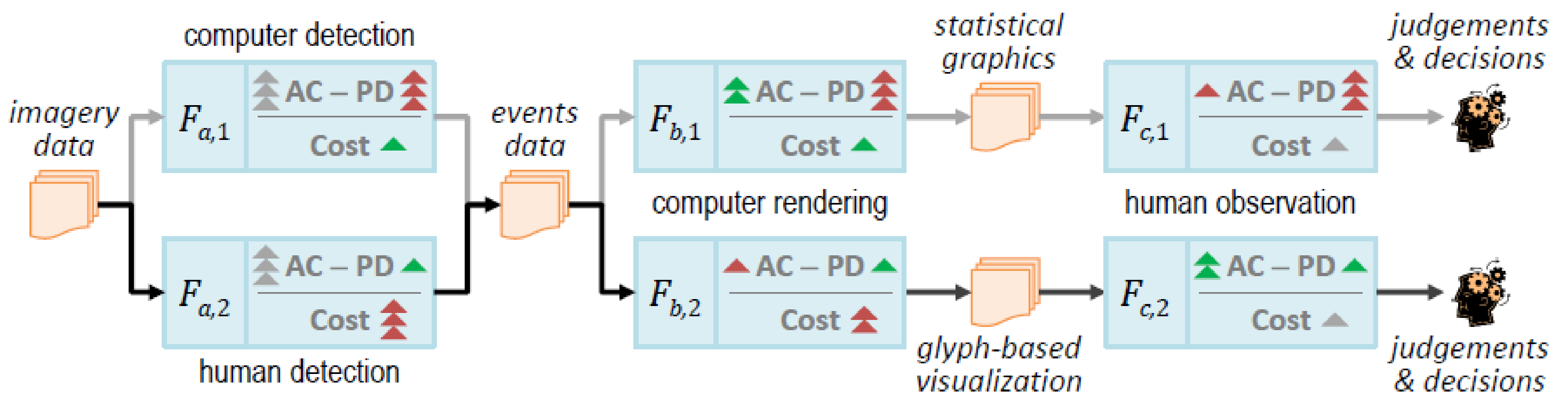
P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths and M. Chen, **MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis,** *Computer Graphics Forum*, 2012

Observational Visualization



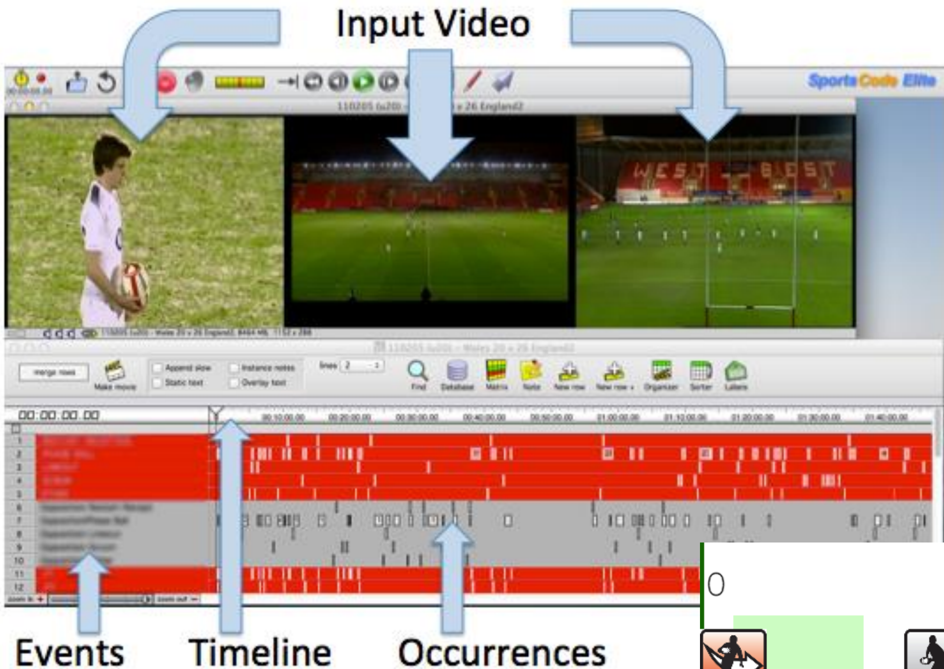
- Real-time or offline annotation results in a huge spreadsheet of events

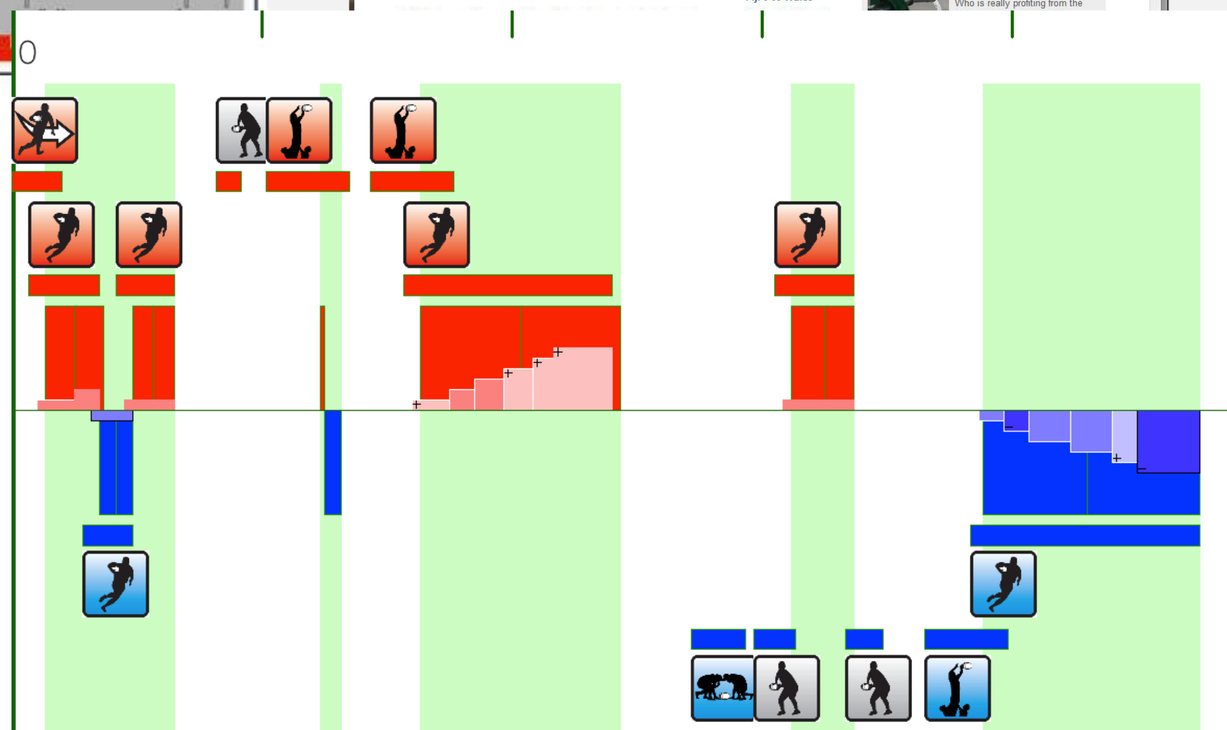




P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths and M. Chen, **MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis,** *Computer Graphics Forum*, 2012

Observational Visualization





Information Theory and Visualization

1. Data Intelligence — a big picture
2. Visualization — a small picture
3. Measurement, Explanation, and Prediction
4. Example: Visual Multiplexing
5. Example: Error Detection and Correction
6. Example: Process Optimization
7. Summary

Information Theory Taxonomy	Relevance in Visualization
Fundamental Concepts	**A possible mathematical framework that underpins the subject of visualization.**
Major Quantities and Properties	Quantitative measurements about the data and visualization space, and the relationship between input and output of a process or subsystem at different stages of a visualization pipeline.
Entropy	Measuring information content (see Section 5.1); salience in visualization.
Mutual Information	Uncertainty reduction in visualization (see Section 5.3); information-assisted visualization.
Major Theorems	Many theorems can be used to explain visualization phenomena and events.
Information balance (conservation law)	Given two visualizations, A and B, the amount of information about A contained in B is the same as that about B in A; overview + detail visualization; multi-view visualization.
Data processing inequality	After visual mapping, the visualization cannot contain more information than the original data (see Section 6.2); Information cannot be recovered after being degraded by some processes or subsystems in a visualization pipeline.
Channel Types	Providing a theoretical basis for classifying visualization subsystems (see Section 6).
Noiseless channel	Not common in practical visualization pipelines (see Section 3).
Noisy channel	Most visualization processes and subsystems can be affected by noise (see Section 3).
Channel Capacity	It can be adapted to define the maximum amount of information that can be visualized or dis-. played (see Sections 5.1 and 6)
Redundancy	Efficiency of a visual mapping; Error detection and correction (see Section 8).
Source Coding (for Noiseless Channels)	**Inspiration for developing new data abstraction and visual encoding techniques.**
Coding Schemes	Applicable, for example, to the following visualization algorithms:
Entropy coding (e.g., Huffman, arithmetic coding)	Logarithmic plots (see Section 7.1); importance-based visualization; information-assisted visualization; magic lens; illustrative deformation.
Dictionary-based coding	Legend design; icon design; visualization literacy; knowledge-assisted visualization.
Run-length encoding	Spatial clustering; cluttering reduction.
Differential encoding	Video visualization; time-varying data visualization.
Subband coding	Multi-resolution modeling; transfer function design in volume rendering.
Transform coding	perceptually-based visual encoding, frequency-domain volume rendering.
Quantization	Color mapping; multi-resolution modeling.
Multiplexing	Comparative visualization; volume rendering; multi-field visualization.