



Eamonn Maguire
CERN

Creating Effective Data Visualizations (a primer)

CSC 2016, September 9th 2016

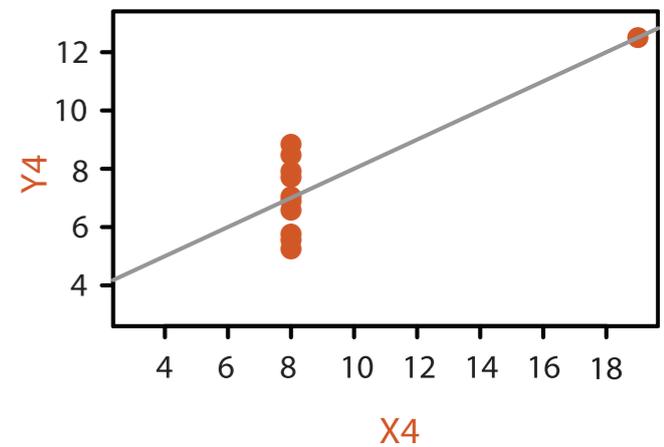
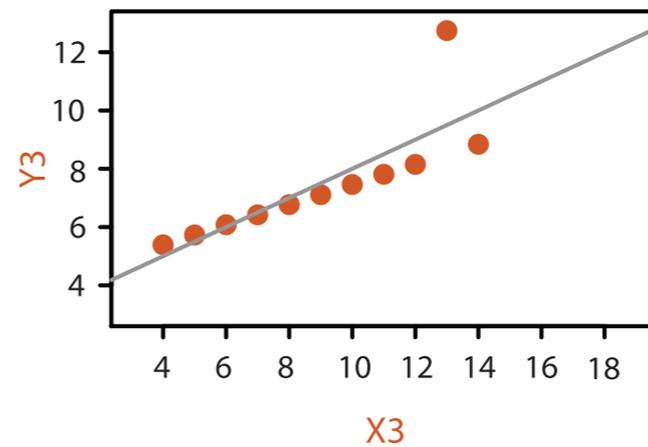
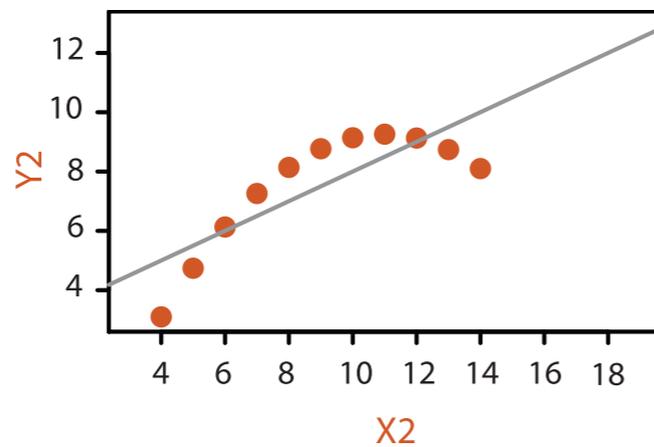
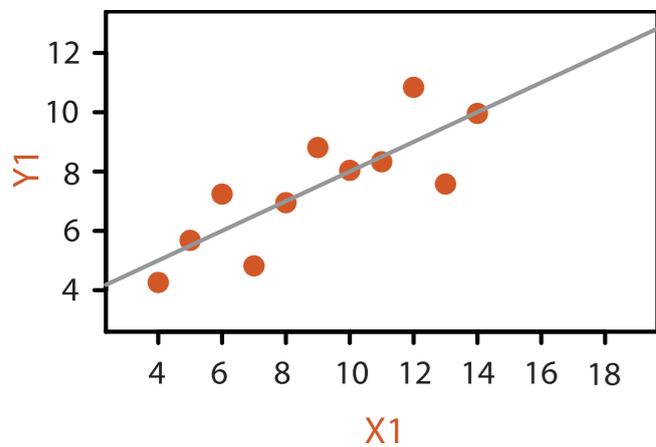
Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

The statistics would lead us to believing that everything is the same



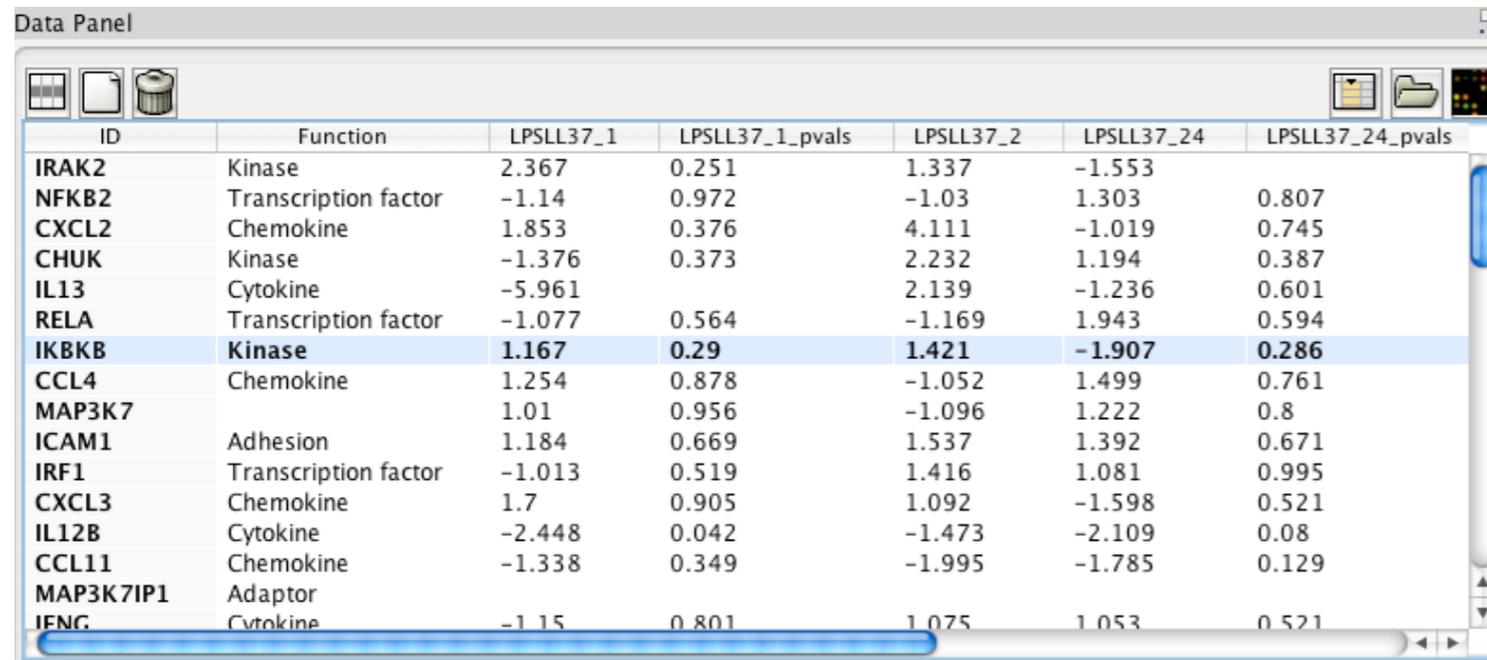
Visualization

The role of visualization systems is to provide visual representations of datasets that help people carry out tasks more effectively.

external representation:
replace cognition with
perception

Visualization

The role of visualization systems is to provide visual representations of datasets that help people carry out tasks more effectively.



ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKBKB	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109	0.08
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785	0.129
MAP3K7IP1	Adaptor					
JENG	Cytokine	-1.15	0.801	1.075	1.053	0.521

external representation:
replace cognition with
perception

Visualization

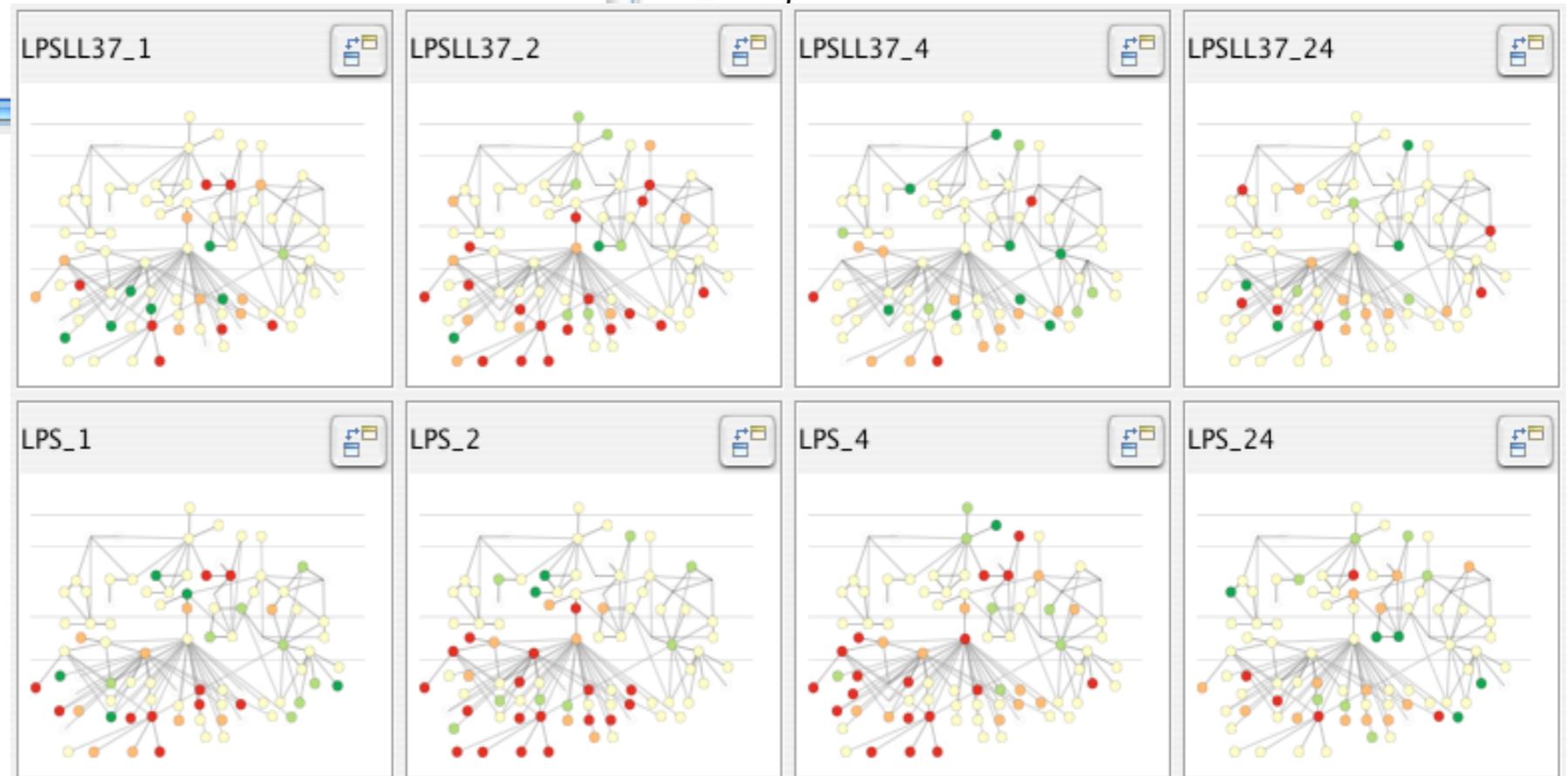
The role of visualization systems is to provide visual representations of datasets that help people carry out tasks more effectively.

Data Panel

ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
RELA	Transcription factor	-1.077	0.564	-1.169	1.943	0.594
IKKBK	Kinase	1.167	0.29	1.421	-1.907	0.286
CCL4	Chemokine	1.254	0.878	-1.052	1.499	0.761
MAP3K7		1.01	0.956	-1.096	1.222	0.8
ICAM1	Adhesion	1.184	0.669	1.537	1.392	0.671
IRF1	Transcription factor	-1.013	0.519	1.416	1.081	0.995
CXCL3	Chemokine	1.7	0.905	1.092	-1.598	0.521
IL12B	Cytokine	-2.448	0.042			
CCL11	Chemokine	-1.338	0.349			
MAP3K7IP1	Adaptor					
JENG	Cytokine	-1.15	0.801			

external representation:
replace cognition with
perception

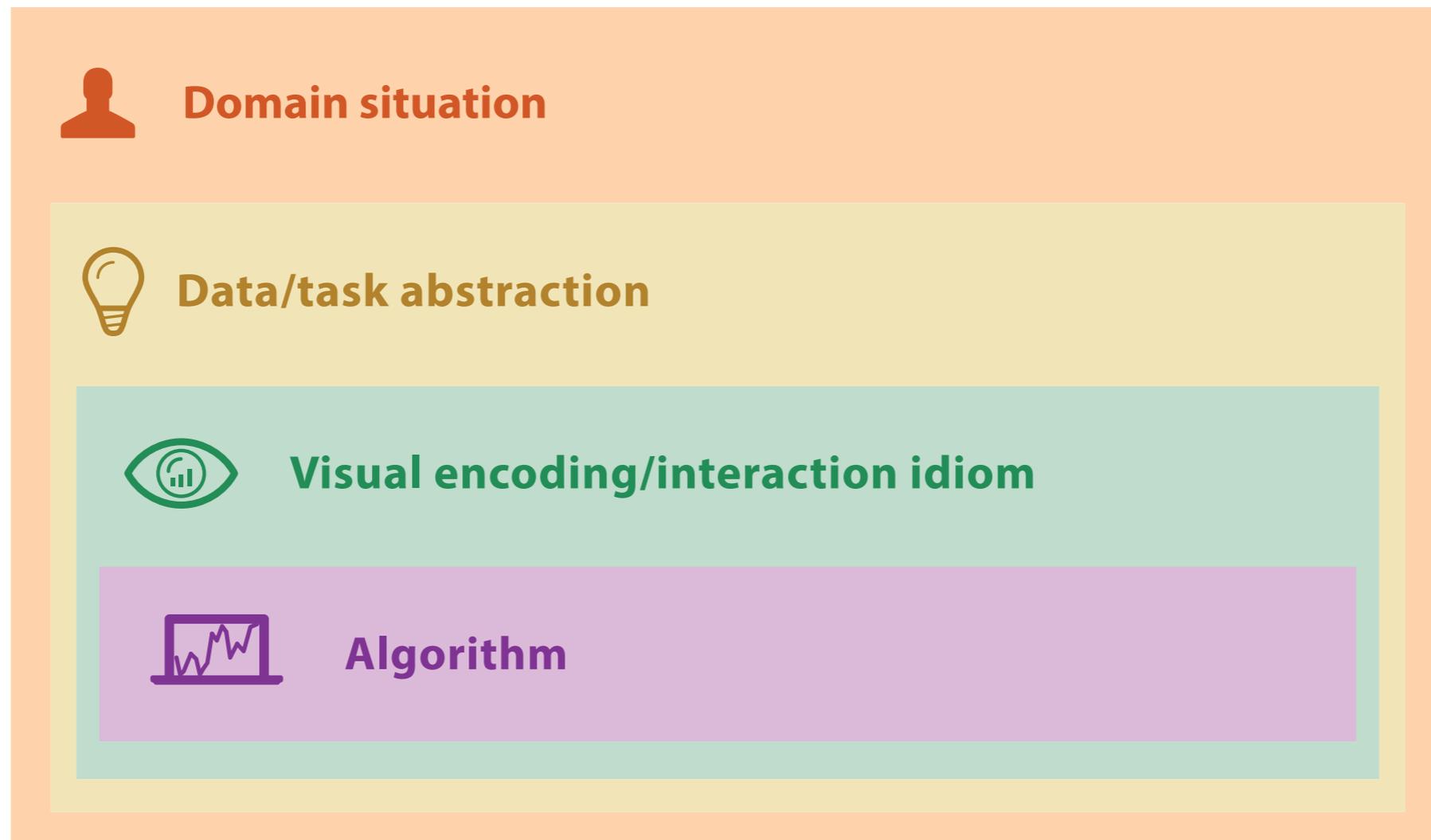
[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. *IEEE TVCG (Proc. InfoVis)* 14(6):1253-1260, 2008.]



How to create effective data visualizations

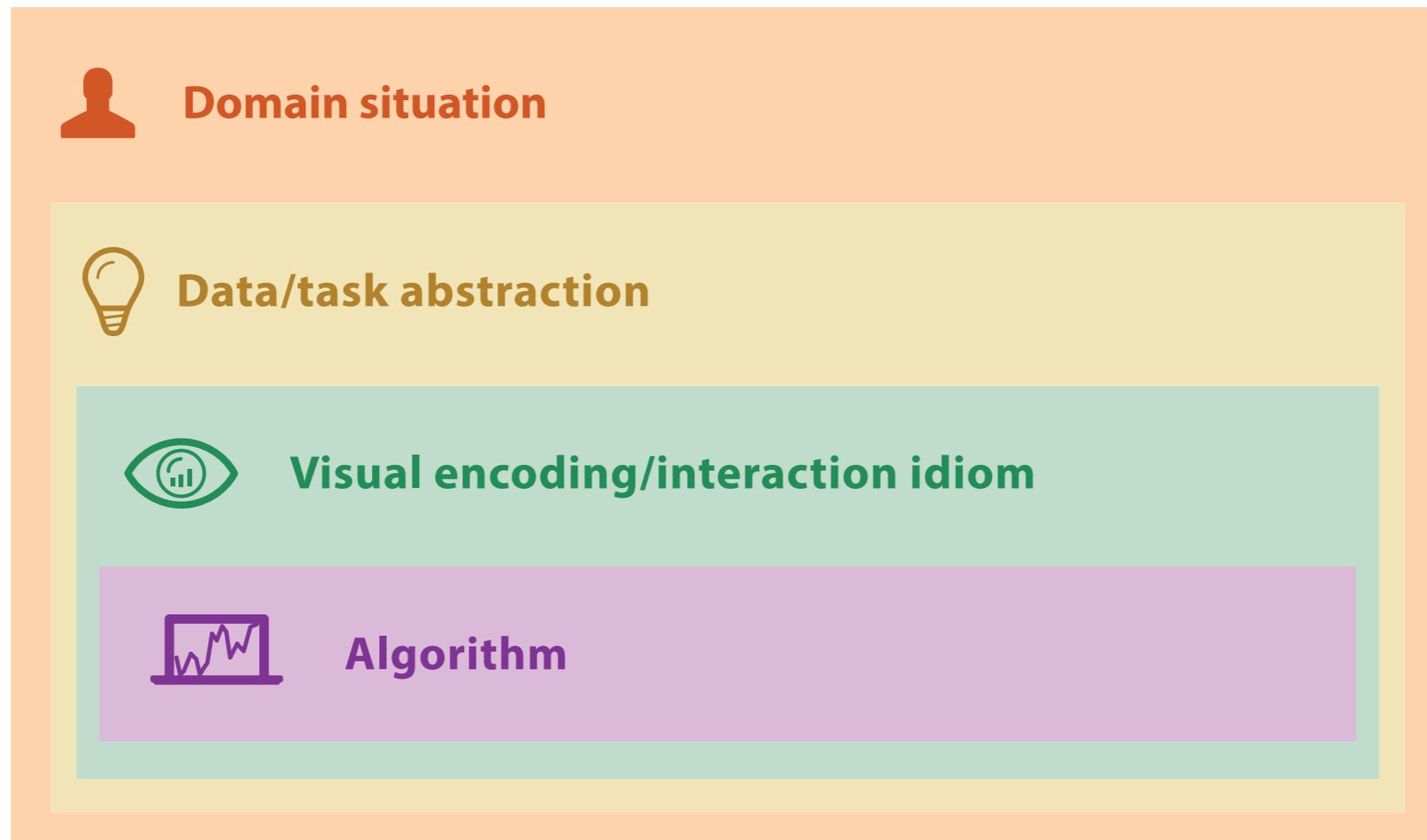
Understand **what** you are visualising,
why you are visualising it,
and **how** you can do it.

How to create effective data visualizations



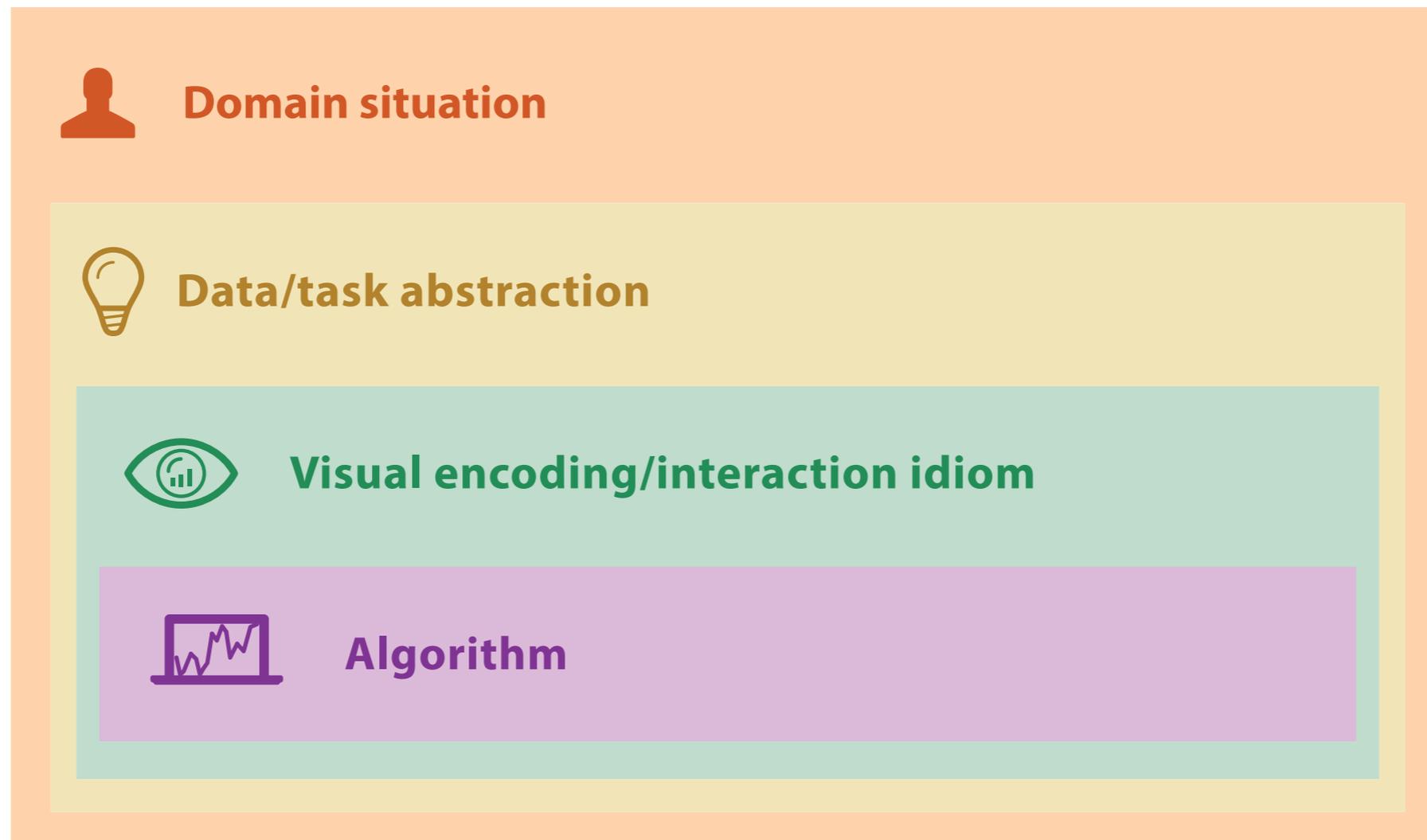
Tamara Munzner, **The Nested Model**

How to create effective data visualizations



what?

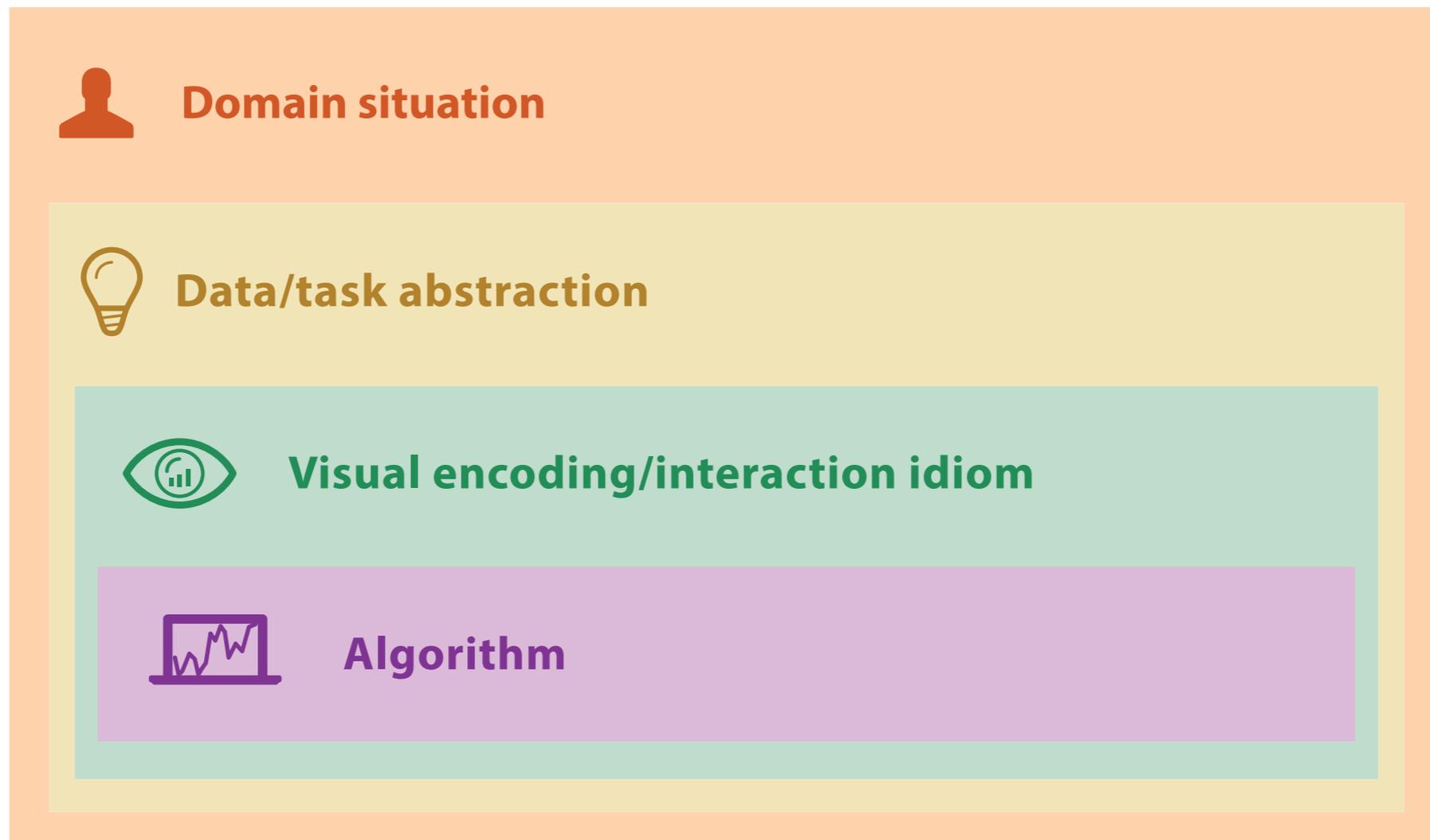
How to create effective data visualizations



what?

why?

How to create effective data visualizations



what?

why?

how?

What are you visualising?

DATA TYPES

→ STATIC

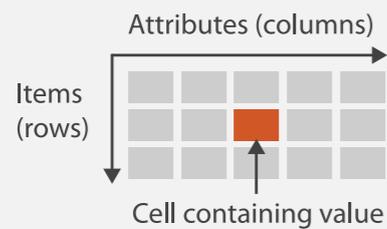


→ DYNAMIC

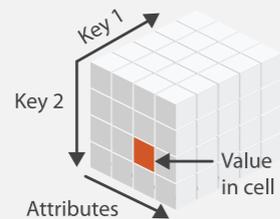


DATASET TYPES

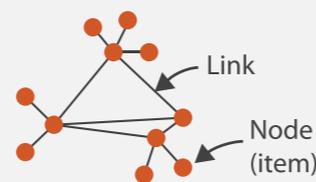
→ TABLES



→ Multidimensional Table



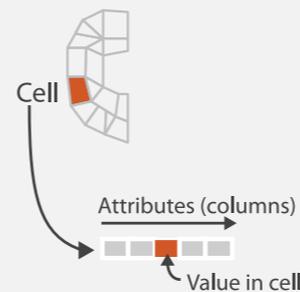
→ NETWORKS



→ Trees



→ FIELDS (CONTINUOUS)



→ GEOMETRY (SPATIAL)



→ TEXT

- Prose Documents
- Document Collections
- Log Files
- Code
- Multimedia

ATTRIBUTE TYPES

→ CATEGORICAL



→ ORDERED

→ Ordinal



→ Quantitative



→ Sequential



→ Diverging



→ Cyclic



Why are you visualising this?

👉 Actions

👉 Use

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



🗄️ Targets

👉 All Data

→ Trends



→ Outliers



→ Features



👉 Search

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

👉 Attributes

→ One

→ Distribution



↓ Extremes



→ Many

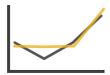
→ Dependency



→ Correlation



→ Similarity



👉 Query

→ Identify



→ Compare



→ Summarise



👉 Network Data

→ Topology

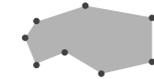


→ Paths



👉 Spatial Data

→ Shape



How can you visualise it?

Encode

Manipulate

Facet

Reduce

Arrange

→ Express



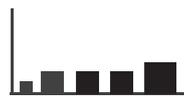
→ Separate



→ Order



→ Align



→ Use



Map

from qualitative and quantitative attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Transparency



→ Region, Texture, Shape, ...



→ Motion

Direction, Rate, Frequency, ...



→ Position, Size, Angle, Curvature, ...



→ Change



→ Select



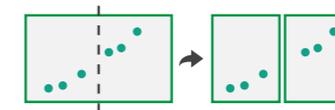
→ Navigate



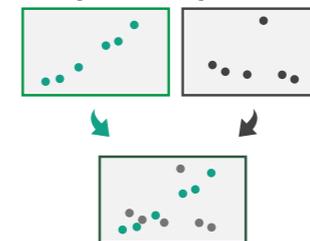
→ Juxtapose



→ Partition



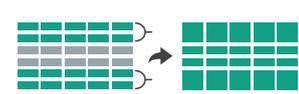
→ Superimpose



→ Filter



→ Aggregate



→ Embed

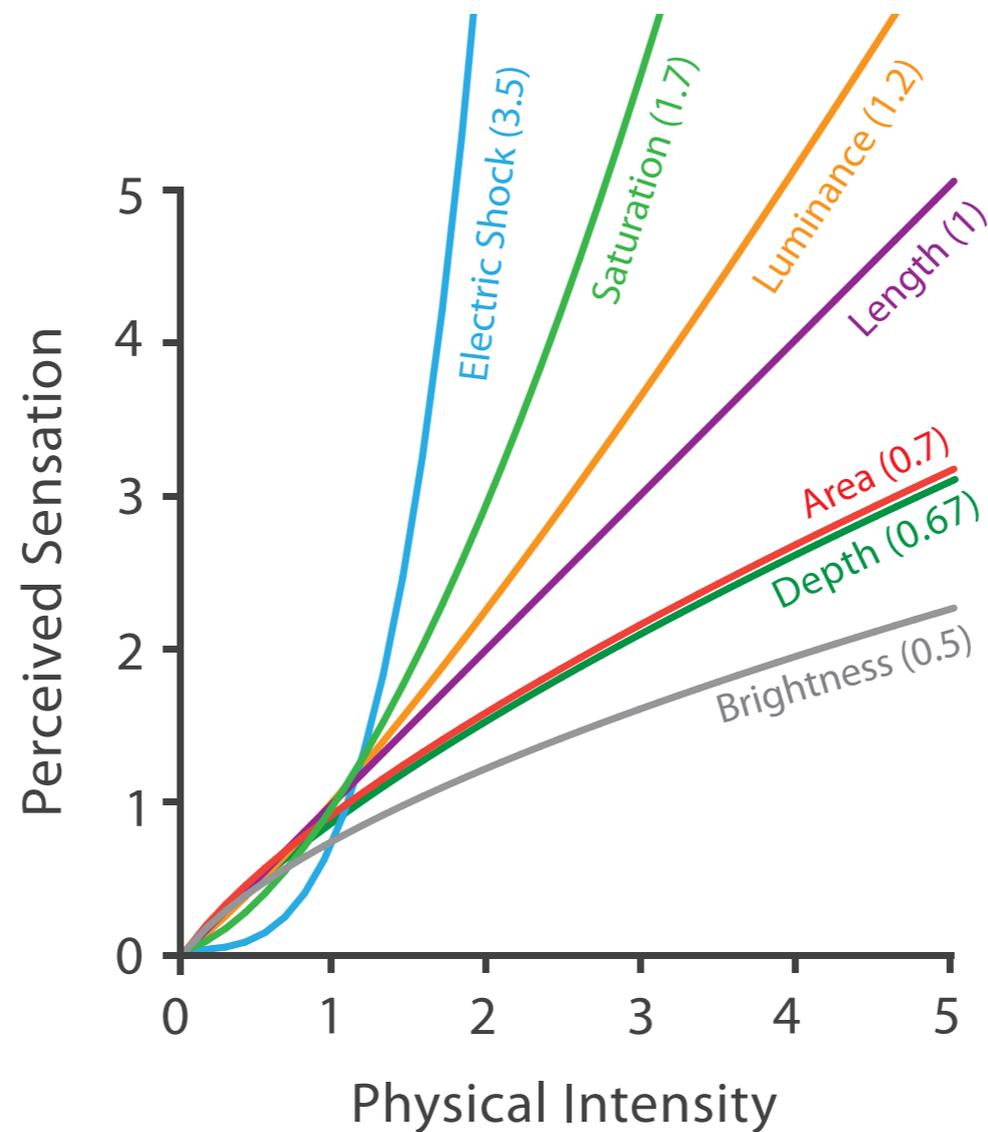


We have to be careful when mapping data to the visual world

Some visual channels are more effective for some data types over others.

Our perception system does not always behave linearly.

Some stimuli are perceived less or more than intended.



Steven's Psychophysical Power Law: $S = I^N$

Stevens, 1975

Quantitative validated

Cleveland and McGill, 1983
Heer and Bostock, 2010
MacKinley, 1986

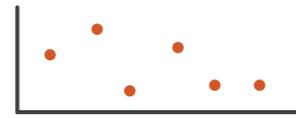
Ordinal not validated

MacKinley, 1986

Categorical not validated

MacKinley, 1986

Suitability of Channel



position (2D)



position (2D)



position (2D)



length (1D size)



texture density



color hue



angle



color saturation



texture pattern



area (2D size)



color hue



connection



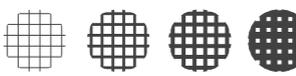
volume (3D size)



texture pattern



containment



texture density



connection



texture density



color saturation



containment



color saturation



color hue



length (1D size)



shape



texture pattern



angle



length (1D size)



connection



area (2D size)



angle



containment



volume (3D size)



area (2D size)



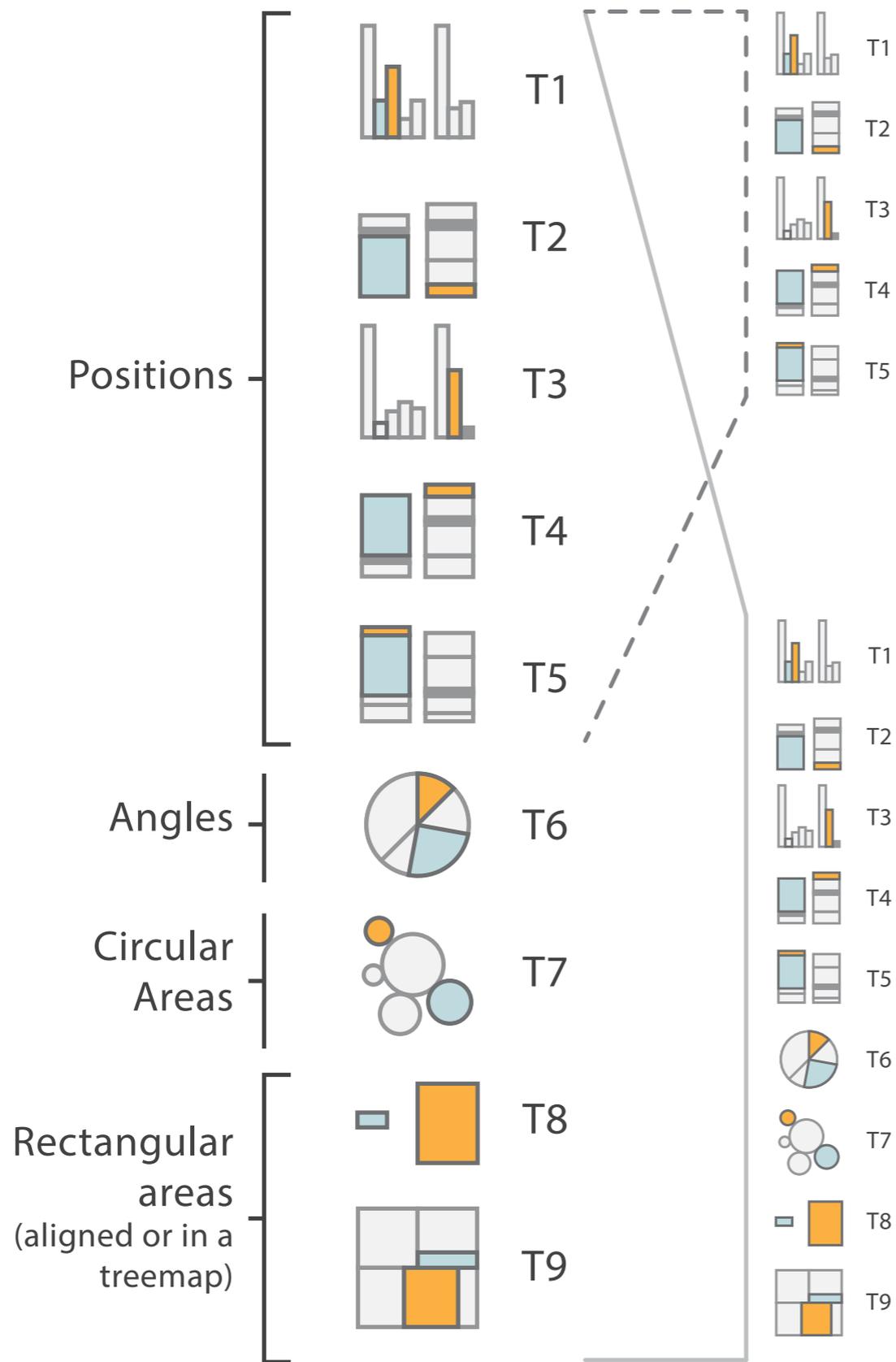
shape



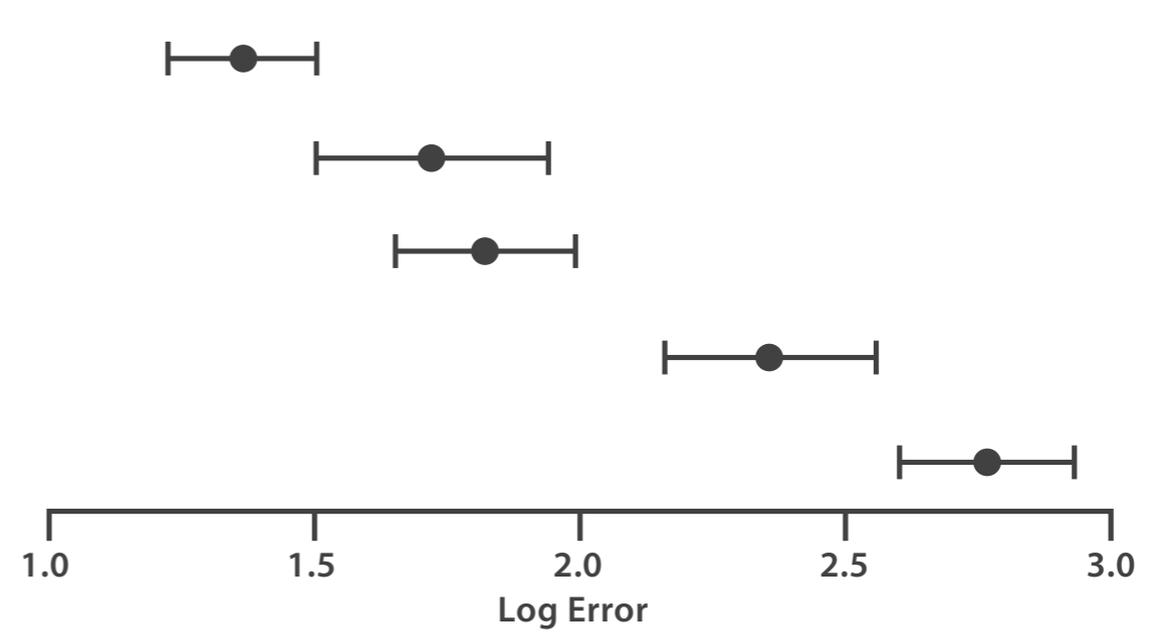
shape



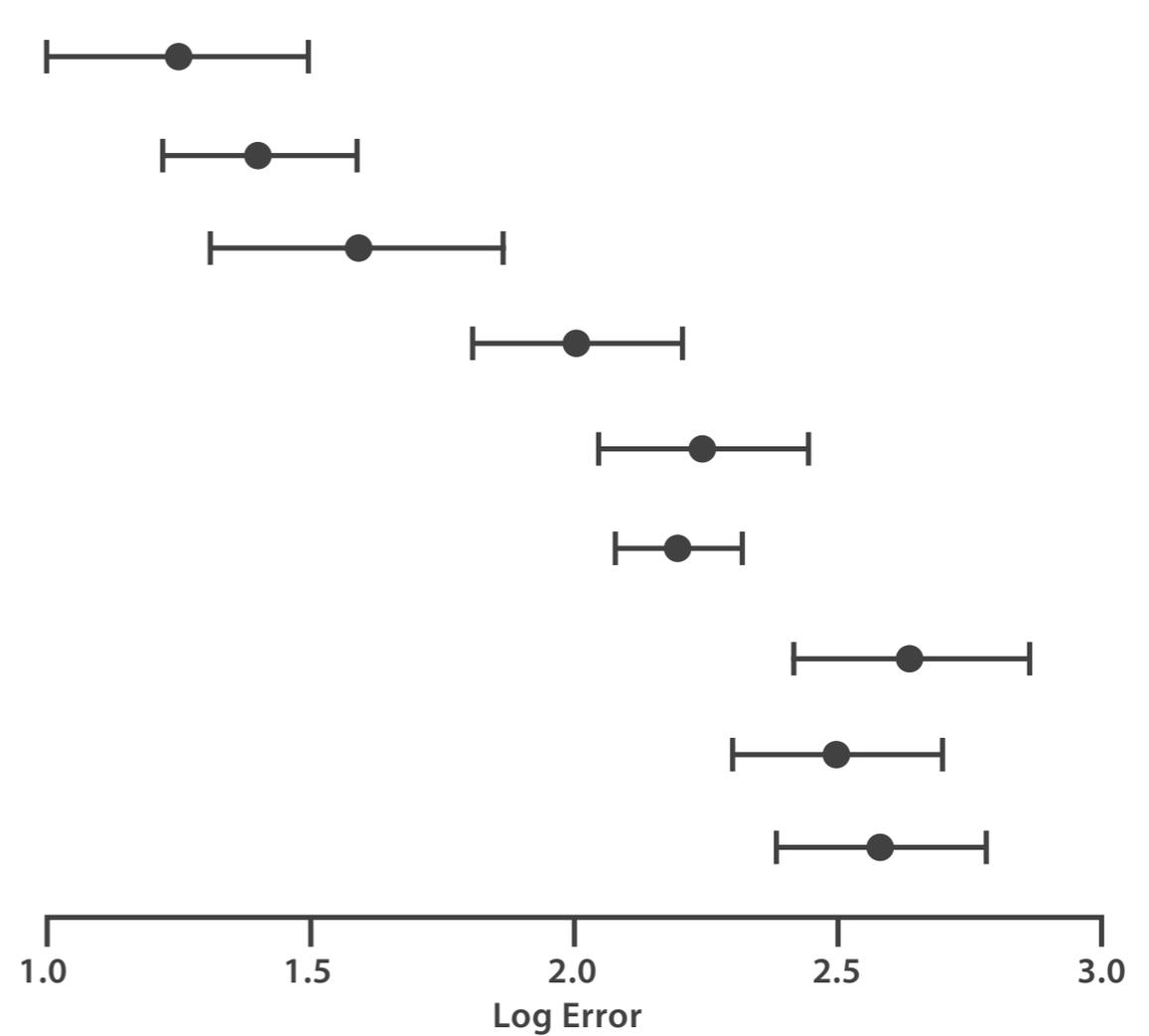
volume (3D size)



Cleveland & McGill's Results (1984)



Heer and Bostock Crowdsourced Results (2010)



How you formulate elements also matters

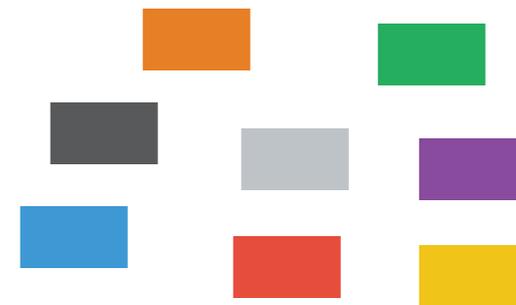
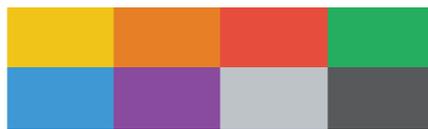
Our visual system makes relative comparisons (except for position)

Limiting what we can do with colour

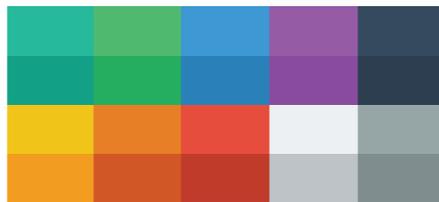


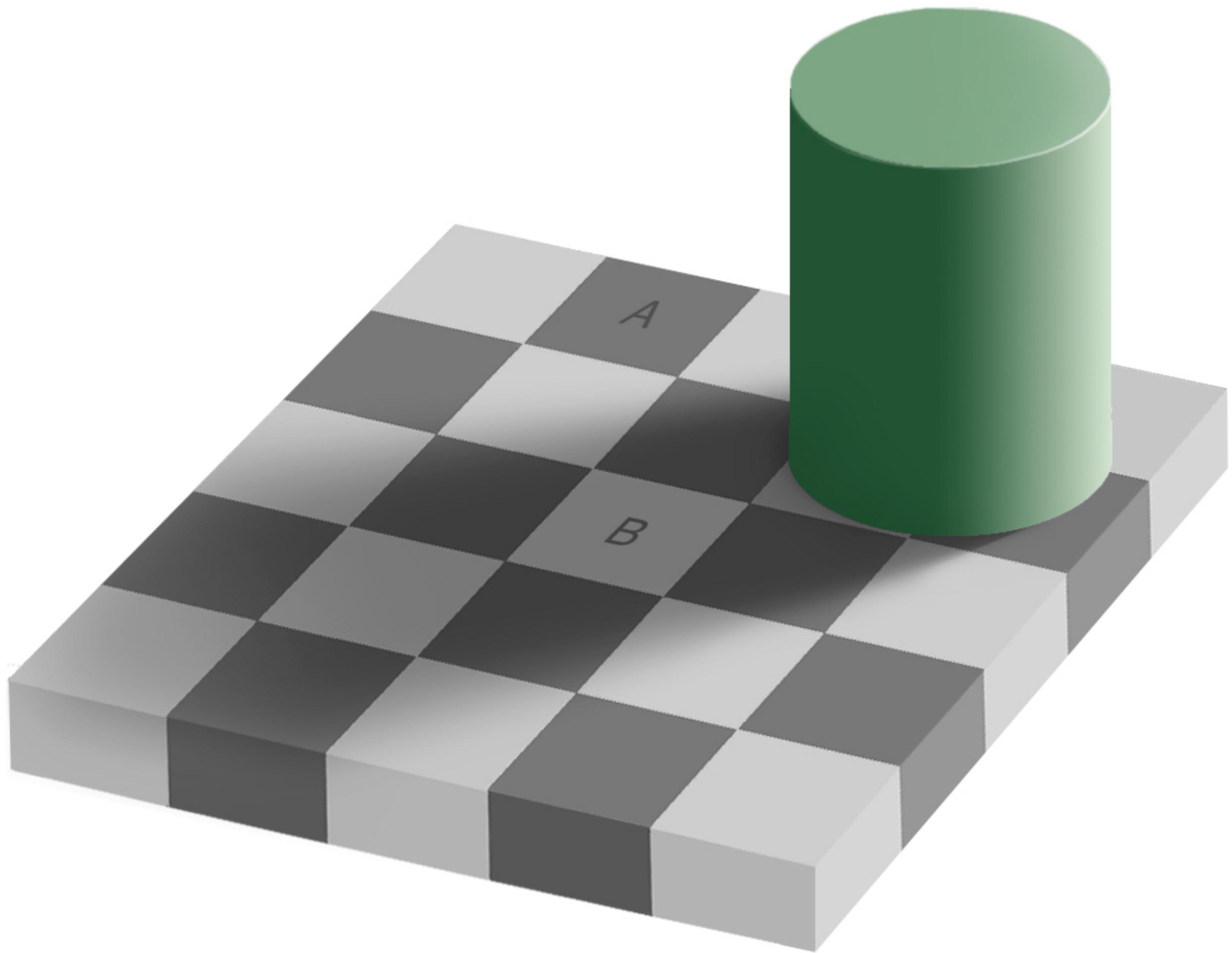
Limiting what we can do with colour

8 values

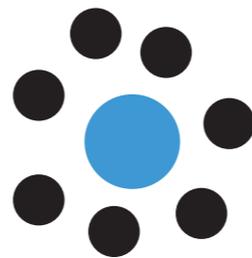
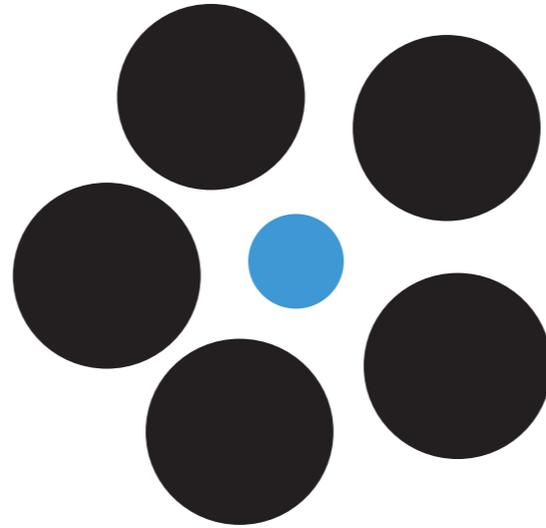


20 values

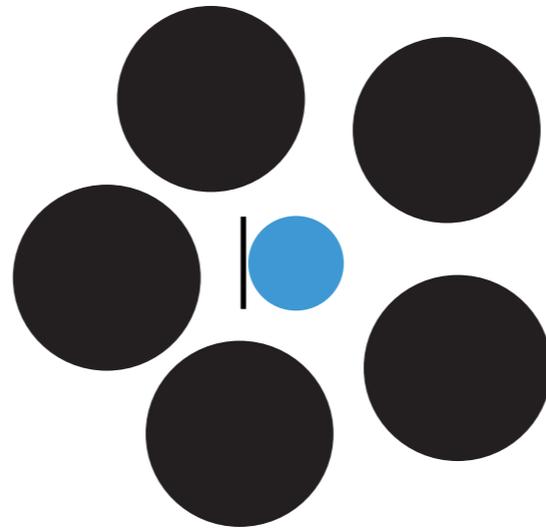




Obscuring how we interpret size



Obscuring how we interpret size



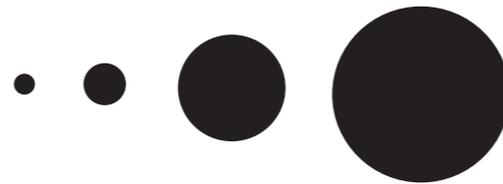
36px |



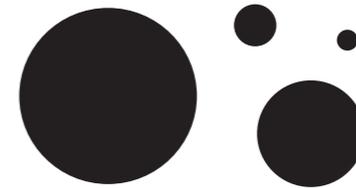
Obscuring how we interpret size

4 values

Aligned



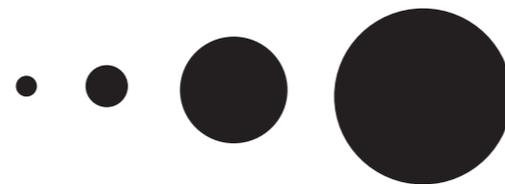
Unordered



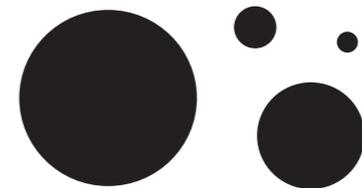
Obscuring how we interpret size

4 values

Aligned

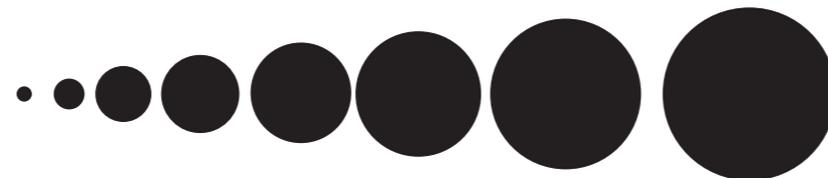


Unordered

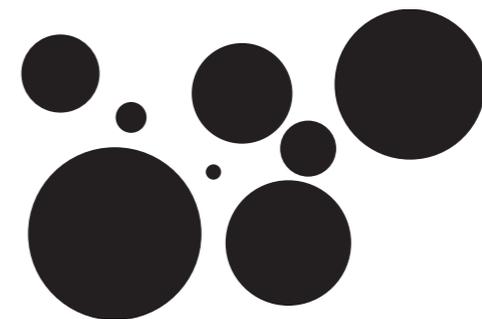


8 values

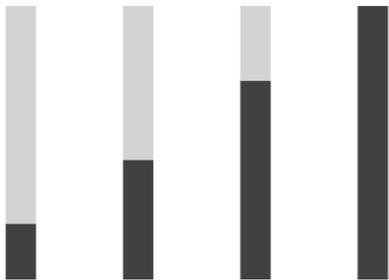
Aligned



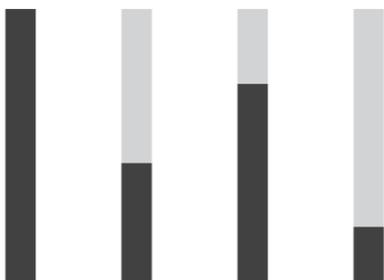
Unordered



Obscuring how we interpret length



4 values



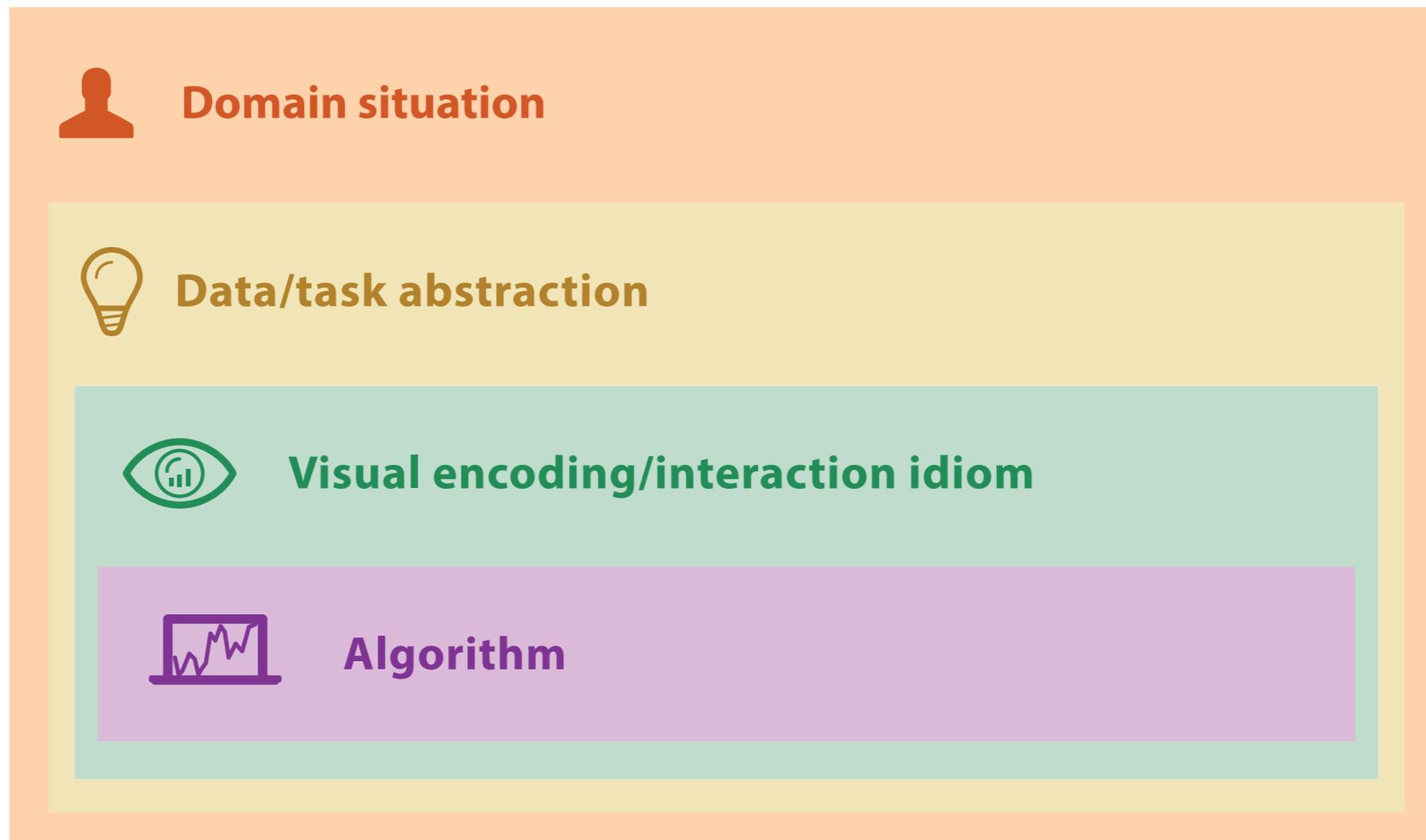
Unordered



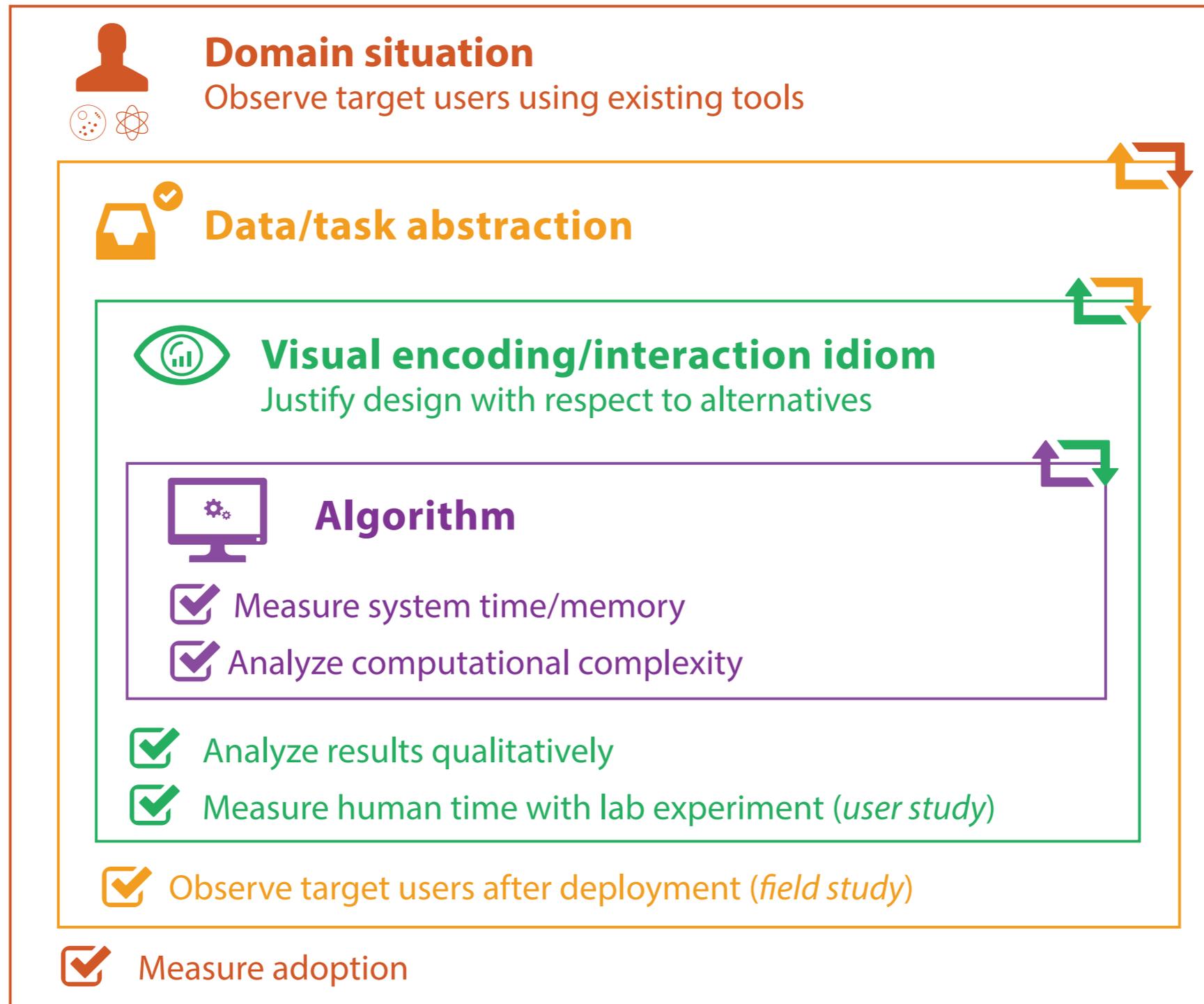
Unaligned

Bringing it all back together

A framework for creating better data visualizations

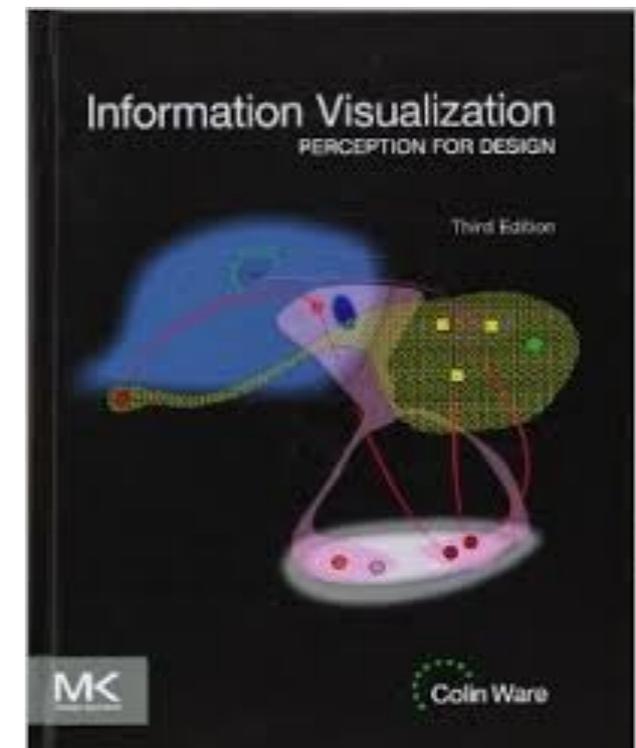
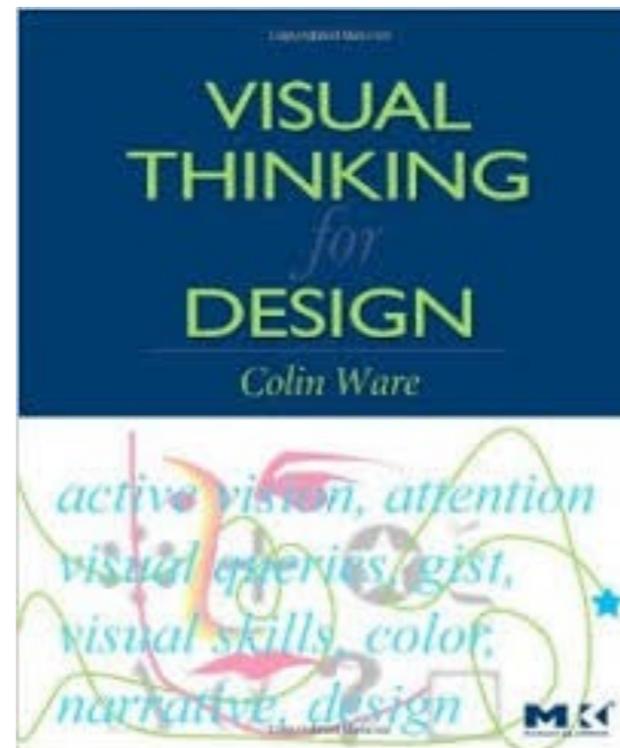
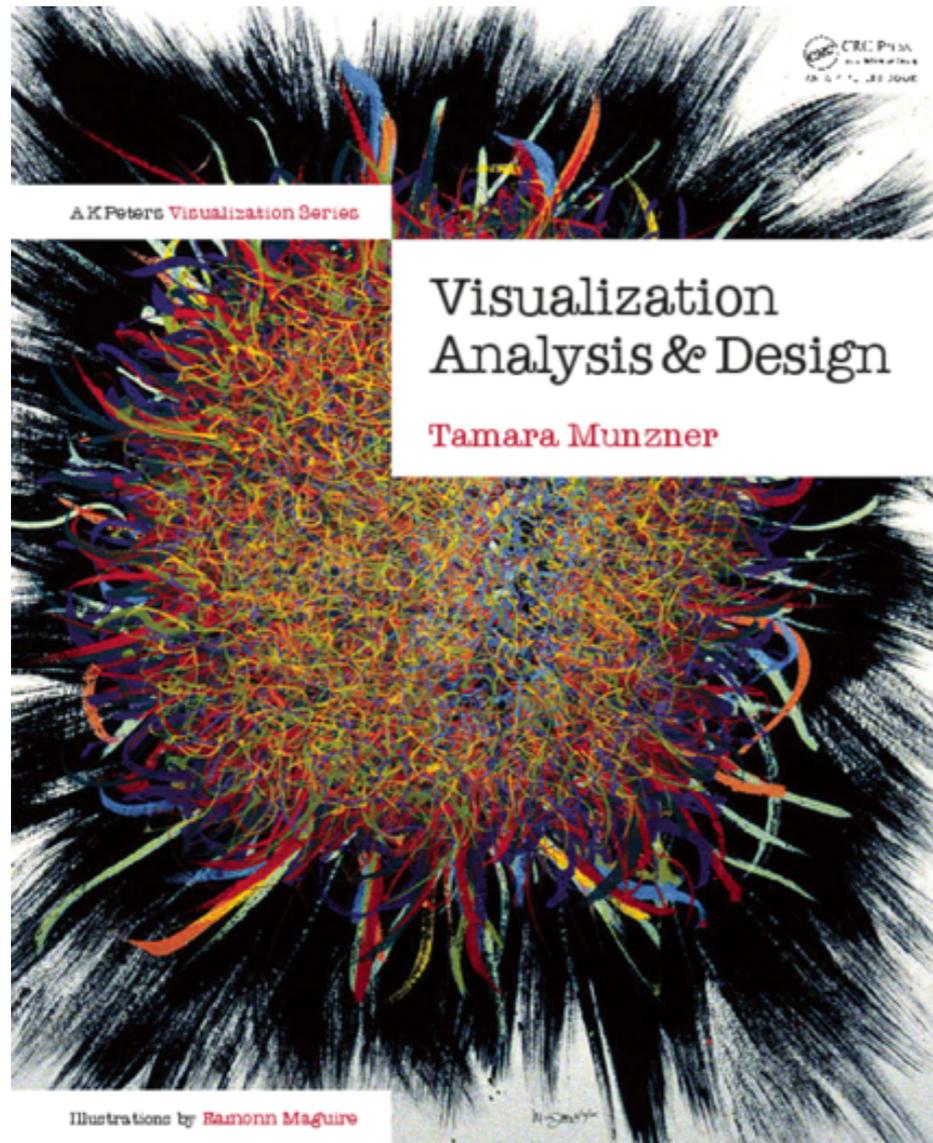


A framework for creating better data visualizations



Tamara Munzner, **The Nested Model**

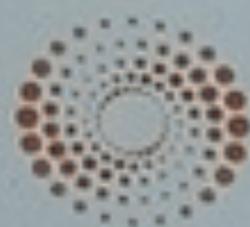
More??



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

Trust
Partnership
Innovation
Performance

**OUR
VALUES**



THOMSON REUTERS™



Questions

[@antarcticdesign](#)
eamonn.maguire@cern.ch

Additional Examples



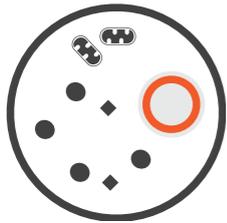
Example

Sequence logo redesign

Primer

From DNA to Protein

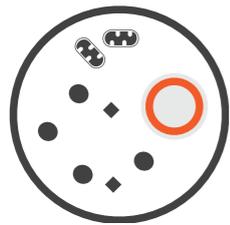
Cell



Primer

From DNA to Protein

Cell



Chromosome



Primer

From DNA to Protein



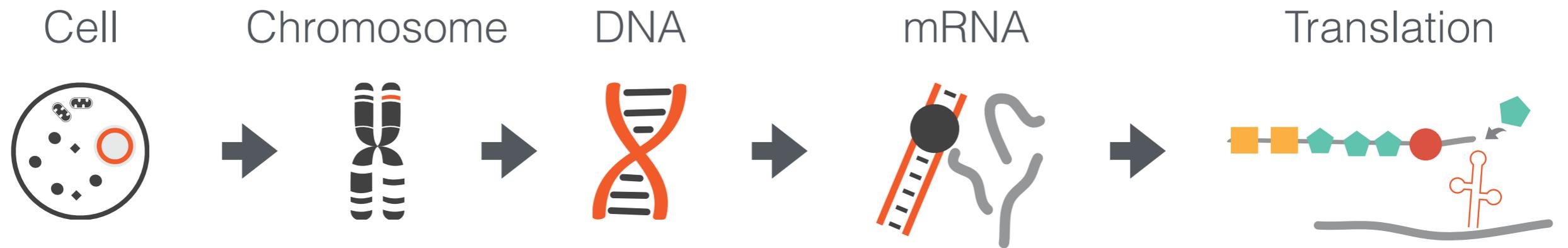
Primer

From DNA to Protein



Primer

From DNA to Protein



Primer

From DNA to Protein



Chromosome



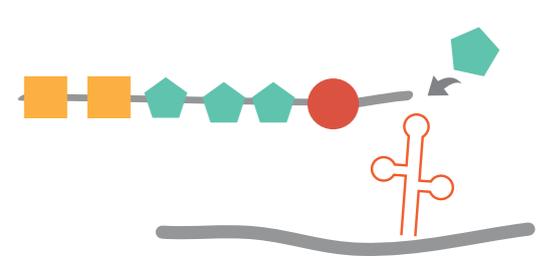
DNA



mRNA



Translation



Peptide Chain

Primer

Fr

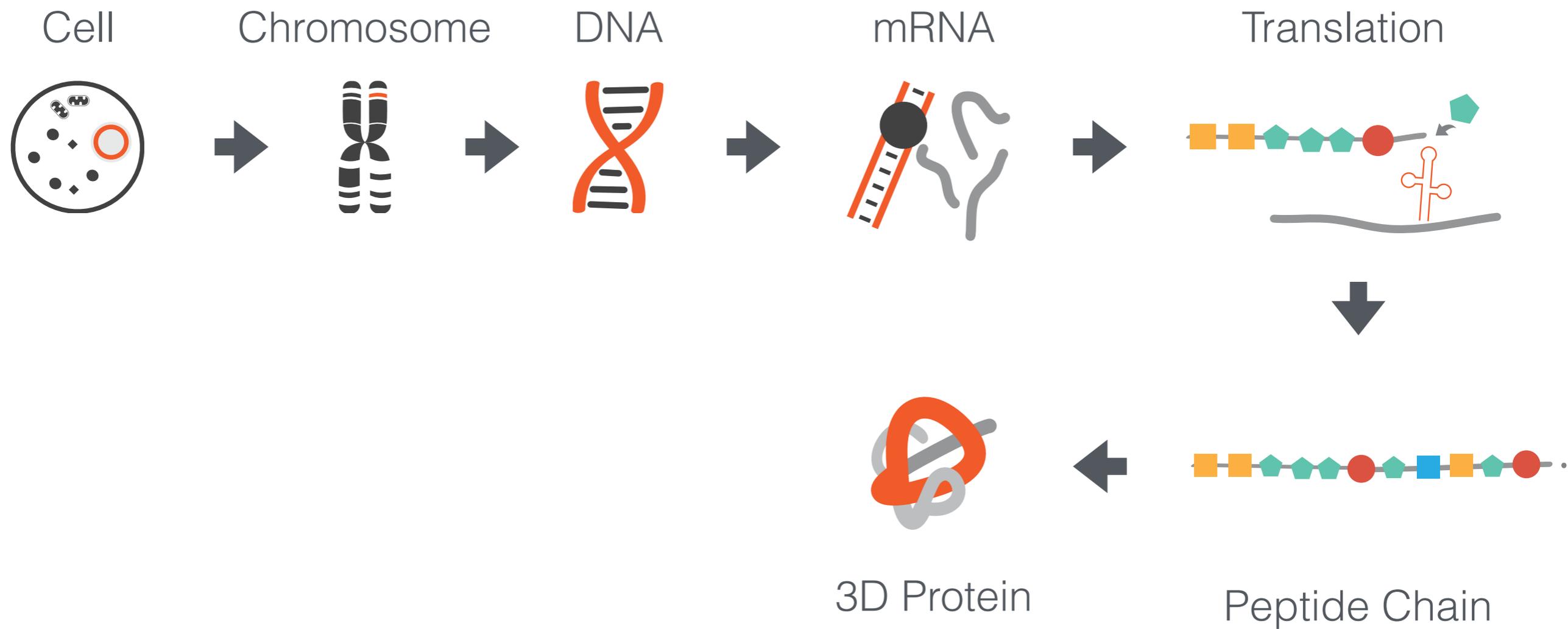
Amino Acid	3-Letter ^[114]	1-Letter ^[114]	Side-chain polarity ^[114]	Side-chain charge (pH 7.4) ^[114]	Hydropathy index ^[115]	Absorbance $\lambda_{\max}(\text{nm})$ ^[116]	ϵ at λ_{\max} ($\times 10^{-3} \text{ M}^{-1} \text{ cm}^{-1}$) ^[116]
Alanine	Ala	A	nonpolar	neutral	1.8		
Arginine	Arg	R	Basic polar	positive	-4.5		
Asparagine	Asn	N	polar	neutral	-3.5		
Aspartic acid	Asp	D	acidic polar	negative	-3.5		
Cysteine	Cys	C	nonpolar	neutral	2.5	250	0.3
Glutamic acid	Glu	E	acidic polar	negative	-3.5		
Glutamine	Gln	Q	polar	neutral	-3.5		
Glycine	Gly	G	nonpolar	neutral	-0.4		
Histidine	His	H	Basic polar	positive(10%) neutral(90%)	-3.2	211	5.9
Isoleucine	Ile	I	nonpolar	neutral	4.5		
Leucine	Leu	L	nonpolar	neutral	3.8		
Lysine	Lys	K	Basic polar	positive	-3.9		
Methionine	Met	M	nonpolar	neutral	1.9		
Phenylalanine	Phe	F	nonpolar	neutral	2.8	257, 206, 188	0.2, 9.3, 60.0
Proline	Pro	P	nonpolar	neutral	-1.6		
Serine	Ser	S	polar	neutral	-0.8		
Threonine	Thr	T	polar	neutral	-0.7		
Tryptophan	Trp	W	nonpolar	neutral	-0.9	280, 219	5.6, 47.0
Tyrosine	Tyr	Y	polar	neutral	-1.3	274, 222, 193	1.4, 8.0, 48.0
Valine	Val	V	nonpolar	neutral	4.2		

From http://en.wikipedia.org/wiki/Amino_acid

in

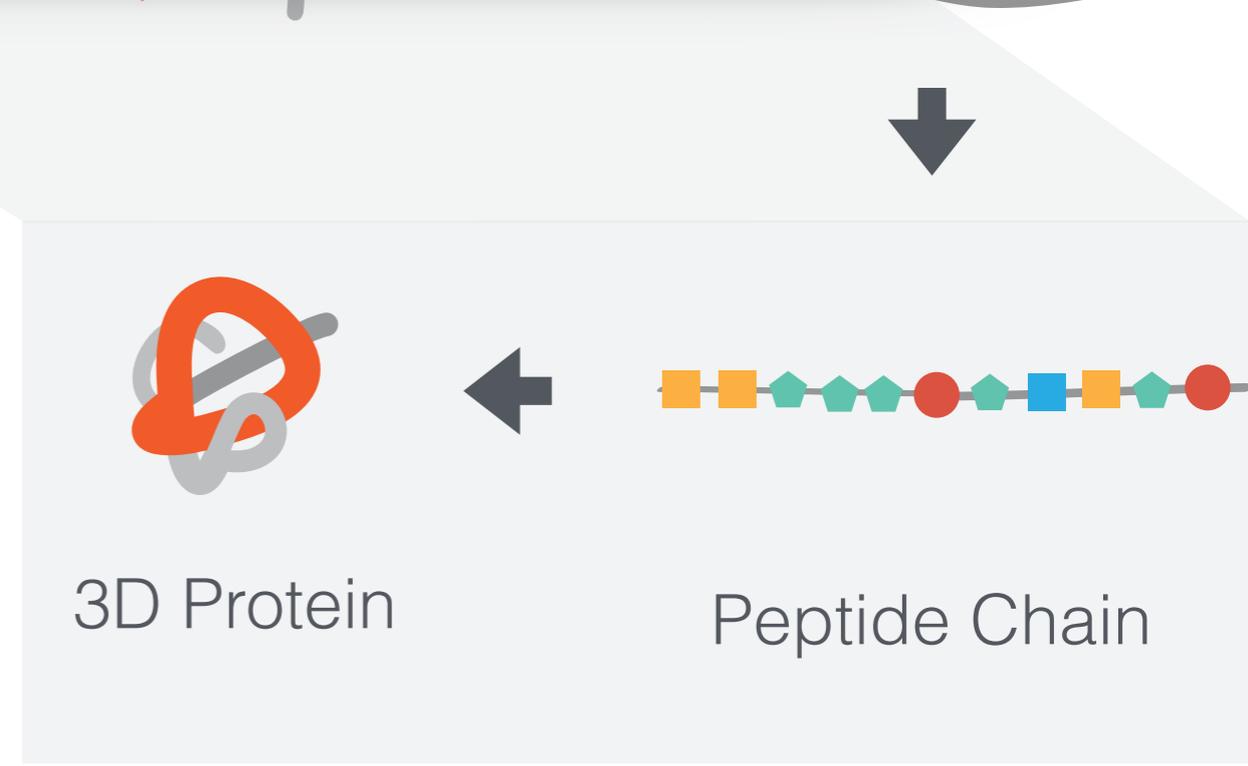
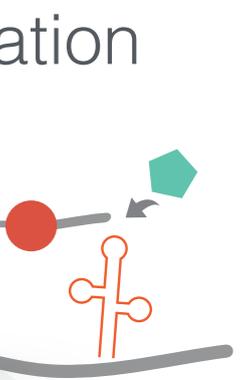
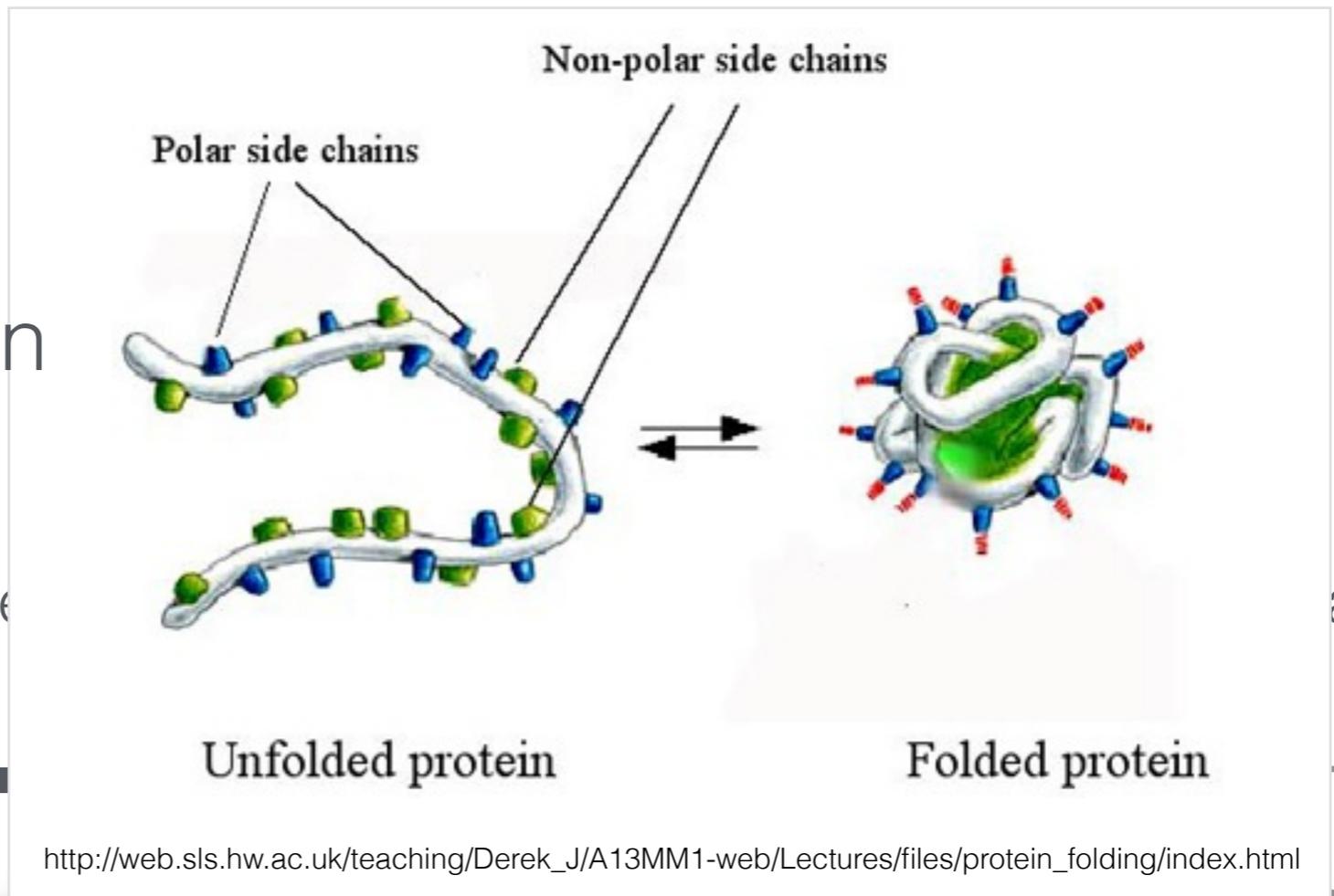
Primer

From DNA to Protein



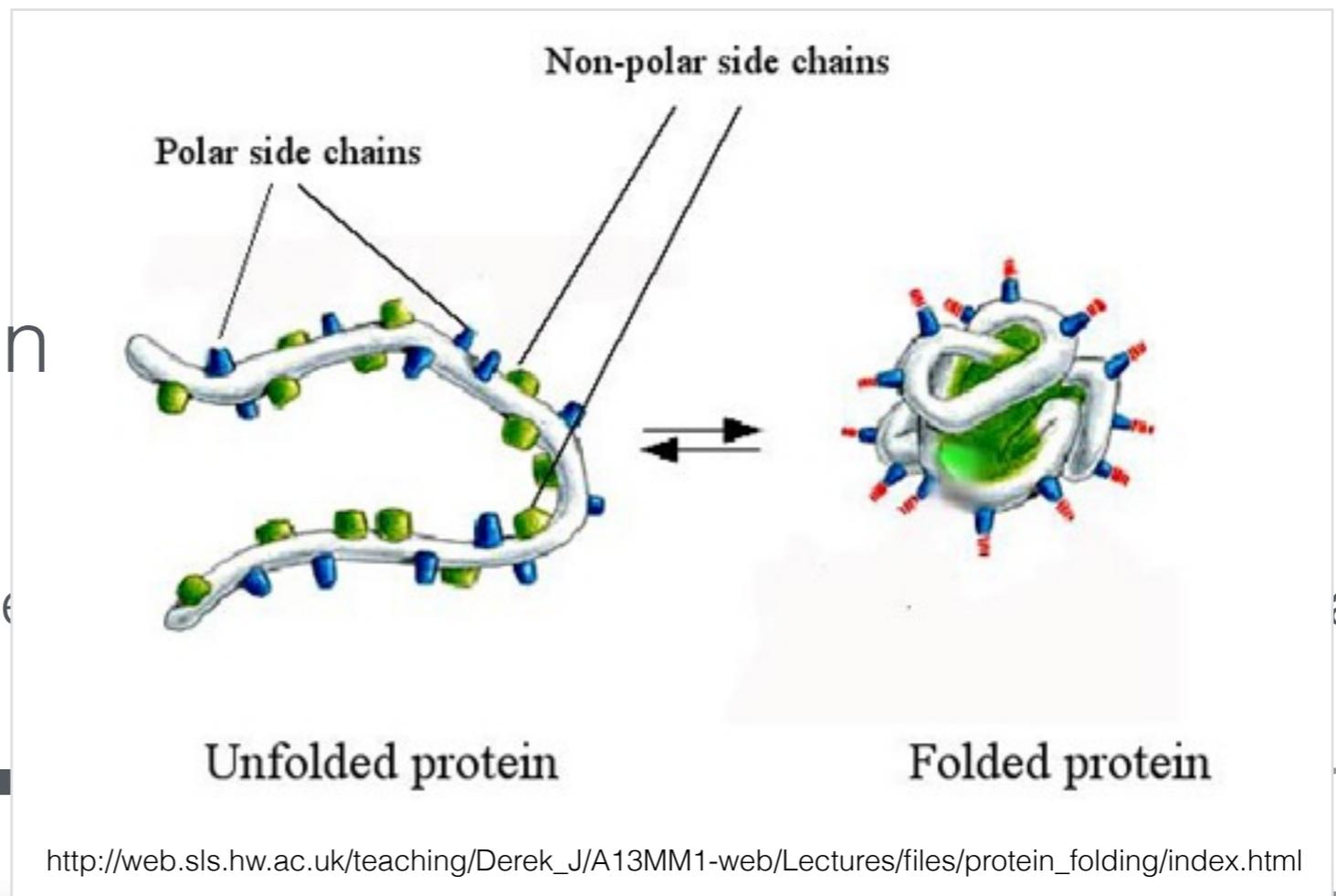
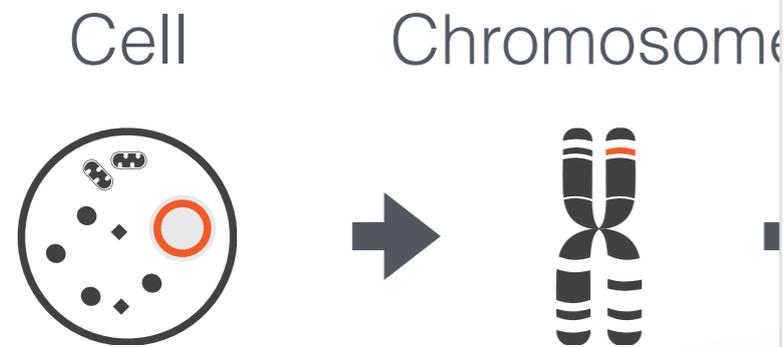
Primer

From DNA to Protein

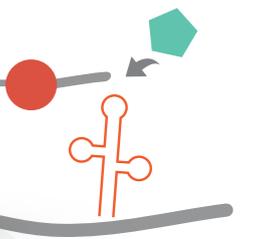


Primer

From DNA to Protein



ation



Protein Structure provides domains that bind to DNA, RNA, other Proteins, and drugs, etc.



3D Protein



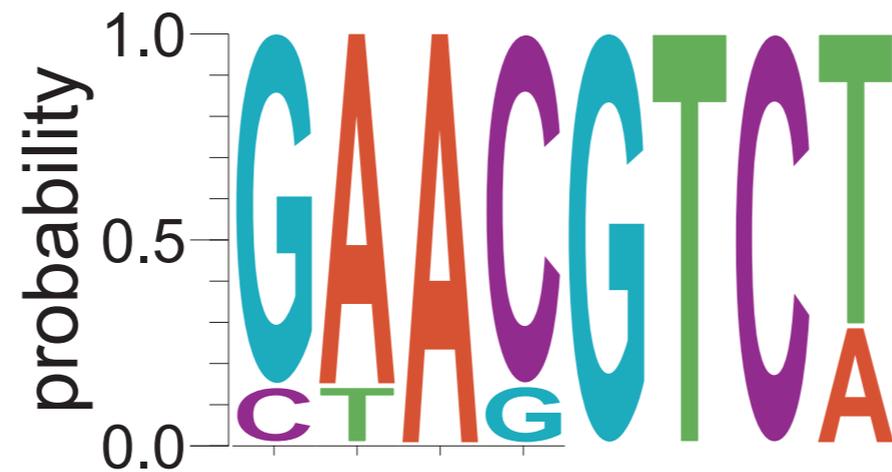
Peptide Chain

What are they?

Sequences



Probability



Entropy



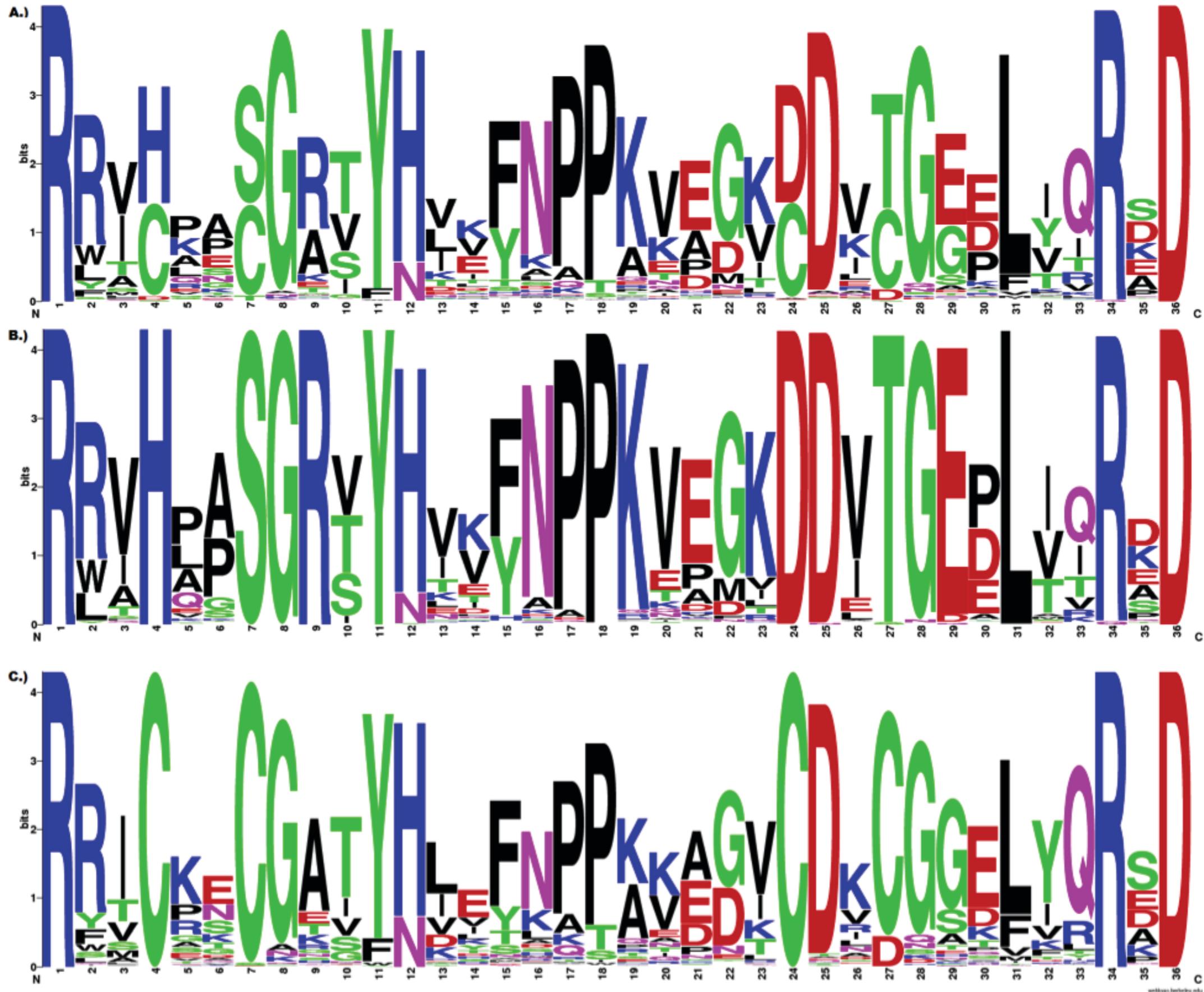
Less uncertainty,
higher bars.

$$\text{Bits} = \text{Log}_2(A)$$

A = Alphabet size

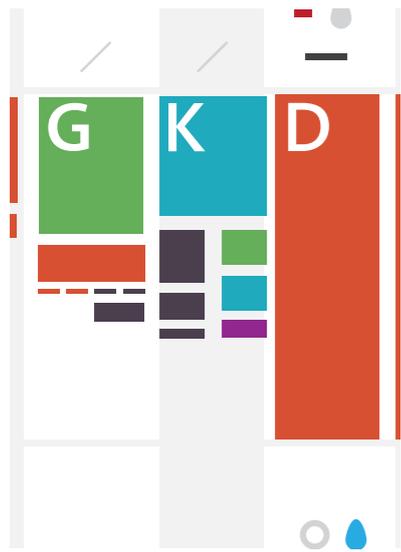
Amino Acids = $\text{Log}_2(20) = 4.32$
DNA/RNA = $\text{Log}_2(4) = 2$

Previous version...

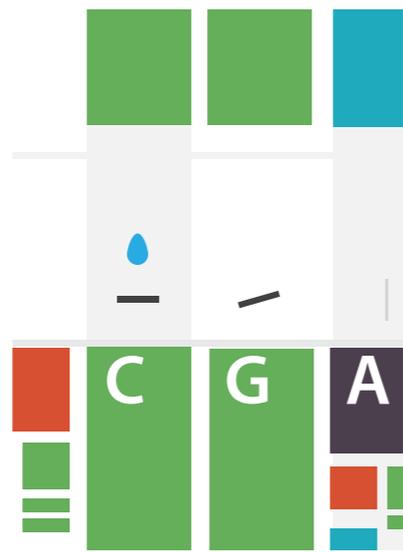


My Solution

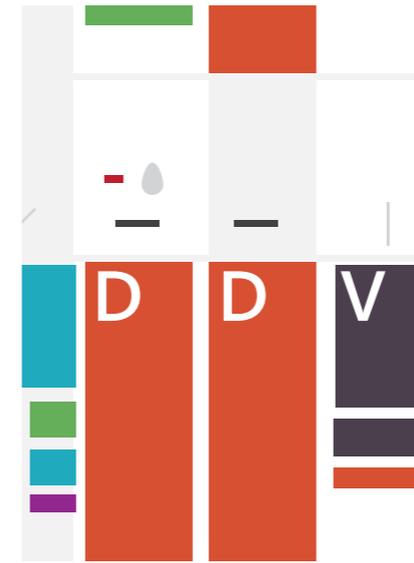
Encoding



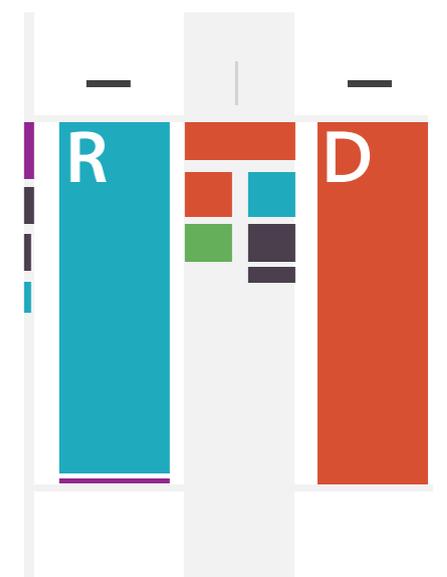
Hydrophobicity



Side-chain Charge

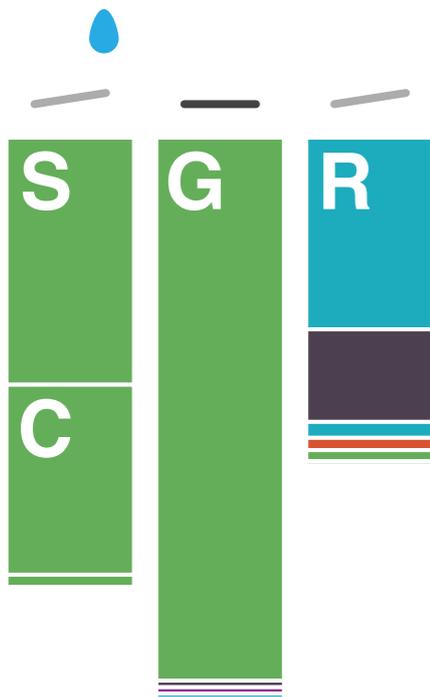
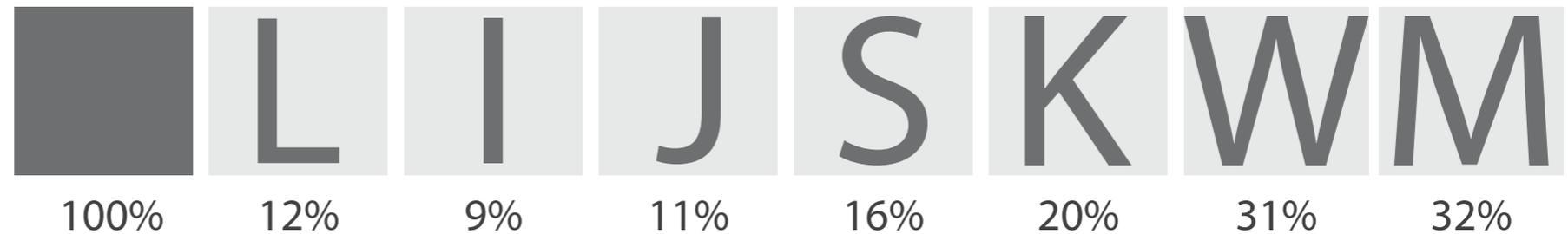


Variation



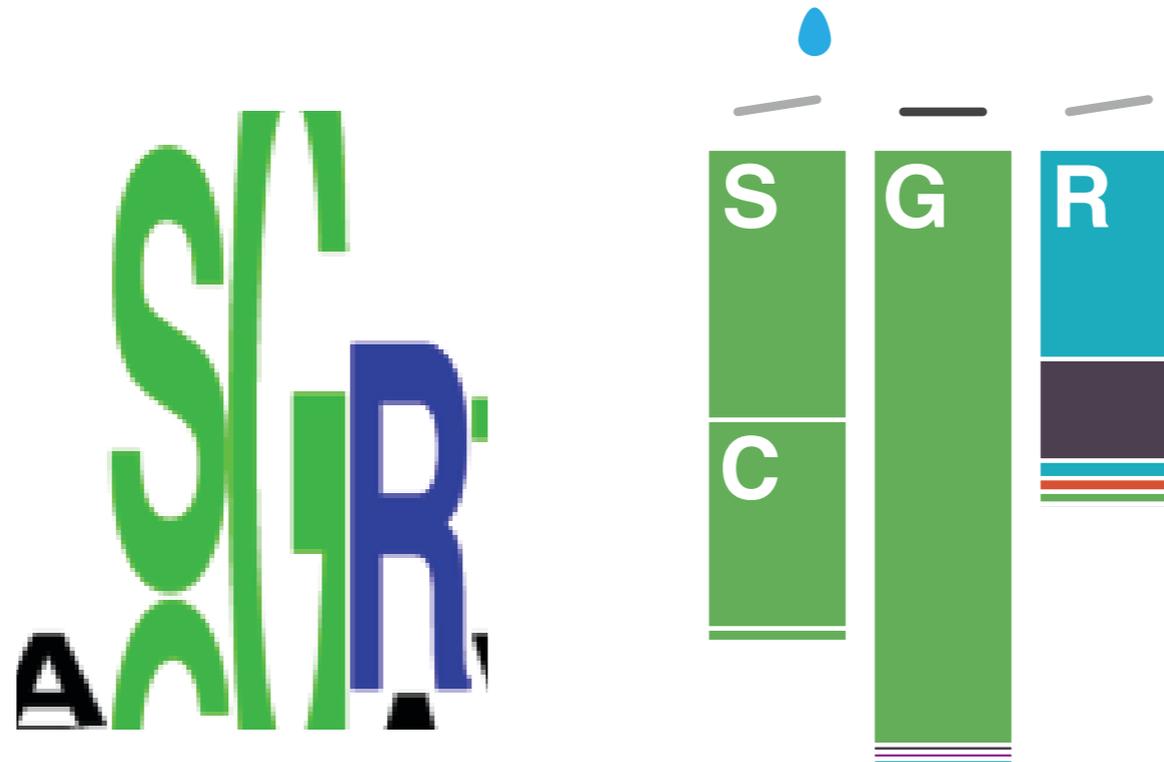
Encoding

Space filling instead of just typography



Using full blocks instead of letters reduces perception problems with the old technique whereby more dense letters will be perceived better than less dense letters, e.g Q versus I.

Encoding

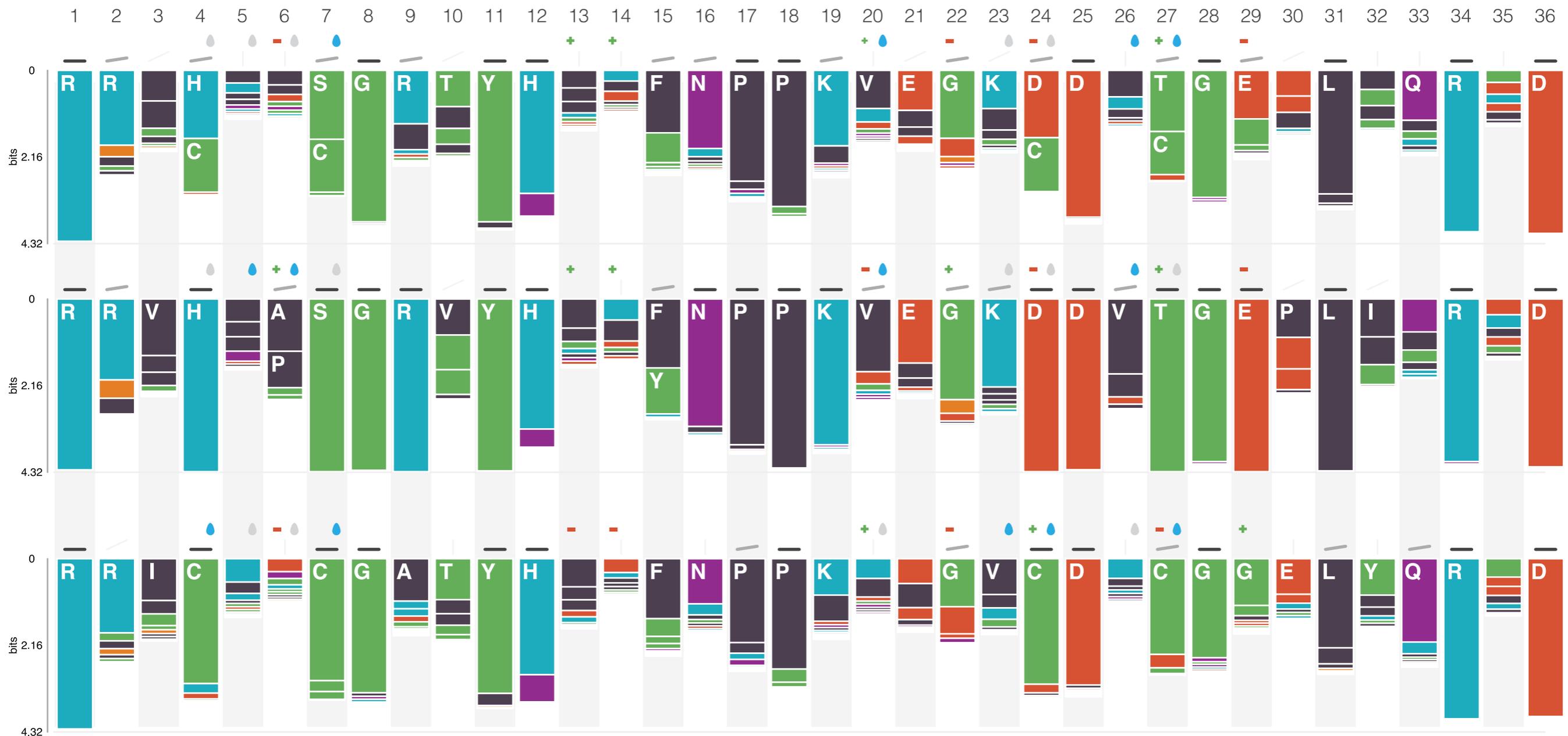


Top amino acids aligned to top to aid better comparison between positions.
e.g. comparing S versus R in the old sequence logo is difficult due to letters
below S making it taller than R.

New version

Evaluated with over 40 domain experts who overwhelmingly preferred the new version, and mis-reads were reduced.

New version



Evaluated with over 40 domain experts who overwhelmingly preferred the new version, and mis-reads were reduced.