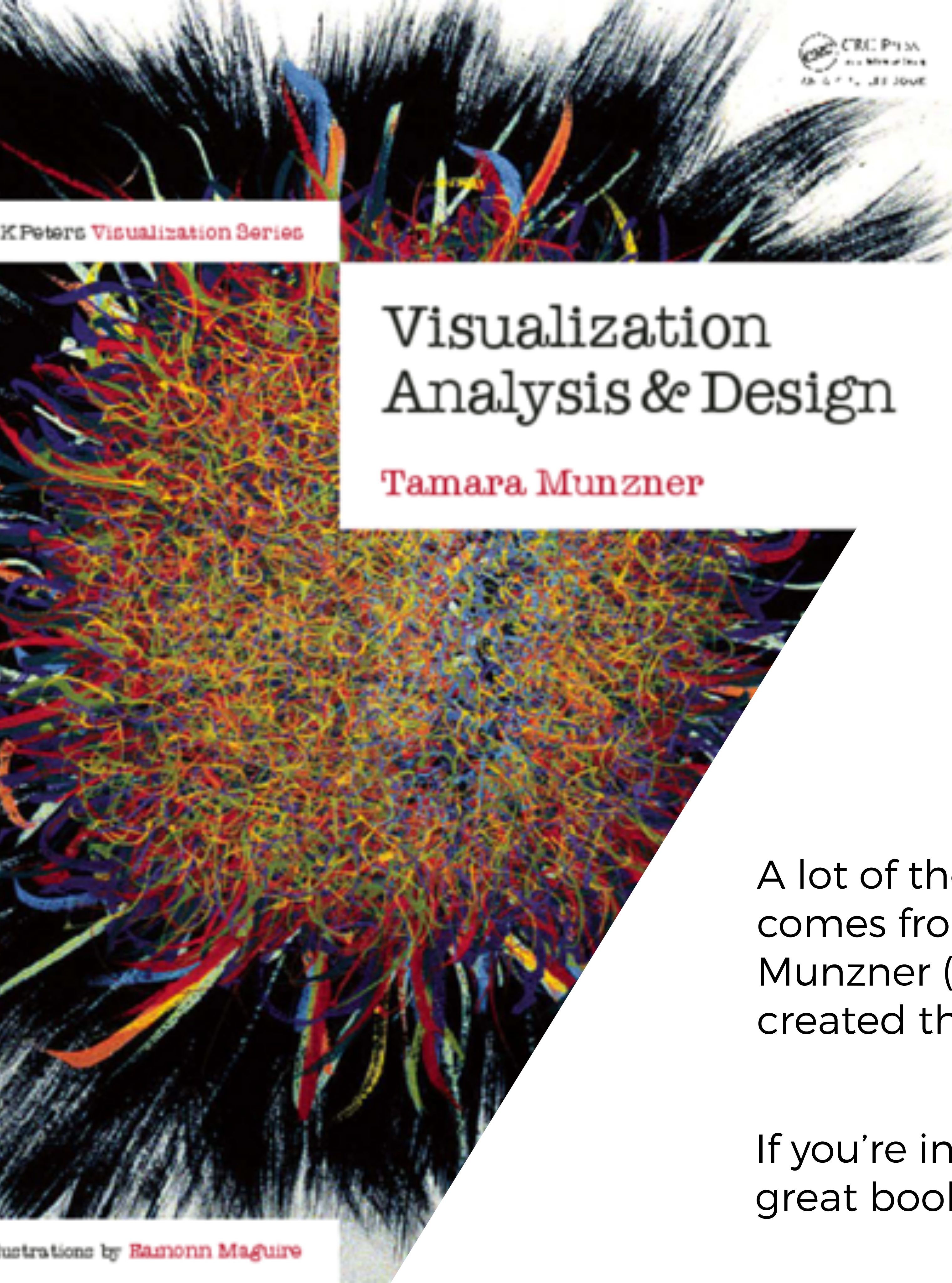
Principles of Data Visualization I

Eamonn Maguire
CERN School of Computing, Romania
September 2019



A lot of the content for this introduction comes from this book from Prof. Tamara Munzner (UBC, Vancouver, Canada) which I created the illustrations for.

If you're interested in learning more, it's a great book to check out :)

Visualization

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

Tamara Munzner

A Visualization should:

1. Save time
2. Have a **clear purpose***
3. Include only the **relevant content***
4. **Encodes data/information** appropriately

* from Noel Illinsky, http://complexdiagrams.com/

Visualization

The role of visualization systems is to provide visual representations of datasets that 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.

Tamara Munzner

A Visualization should:

1. Save time
2. Have a **clear purpose***
3. Include only the **relevant content***
4. **Encodes data/information** appropriately

* from Noel Illinsky, http://complexdiagrams.com/

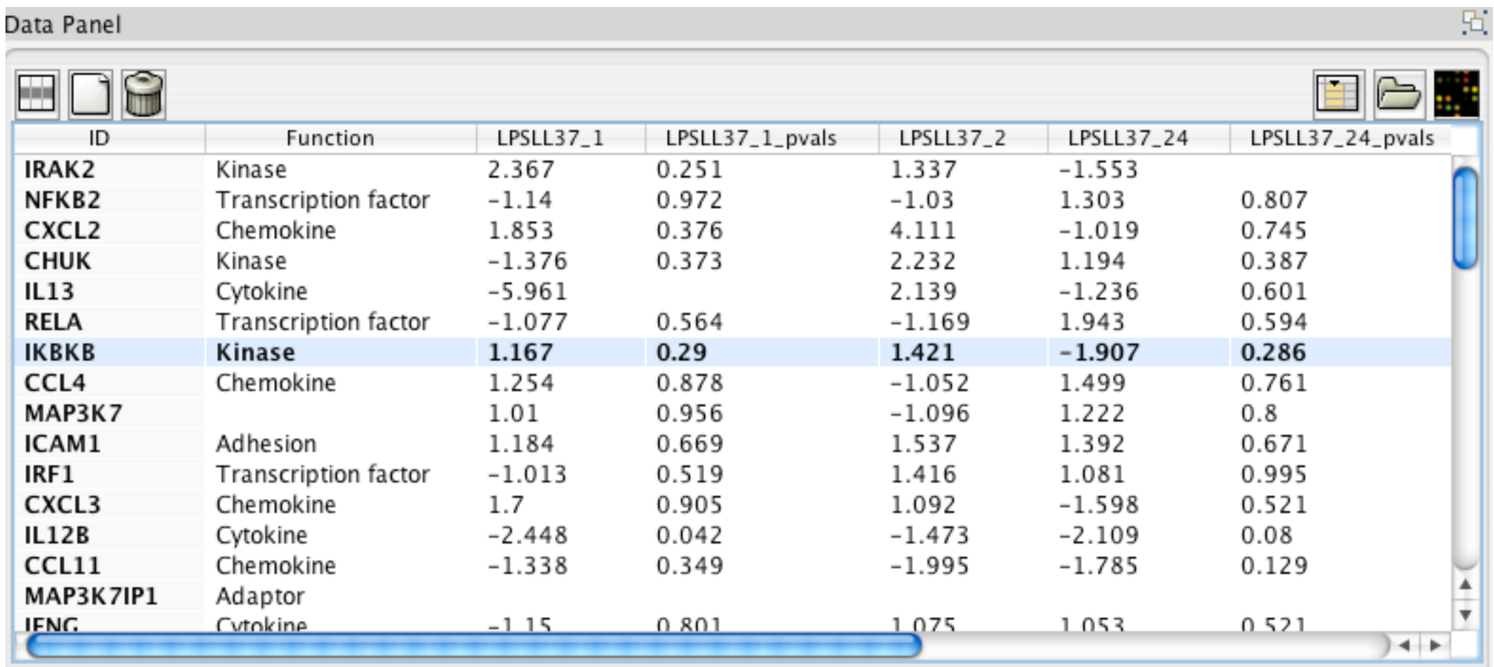
Visualization

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**.



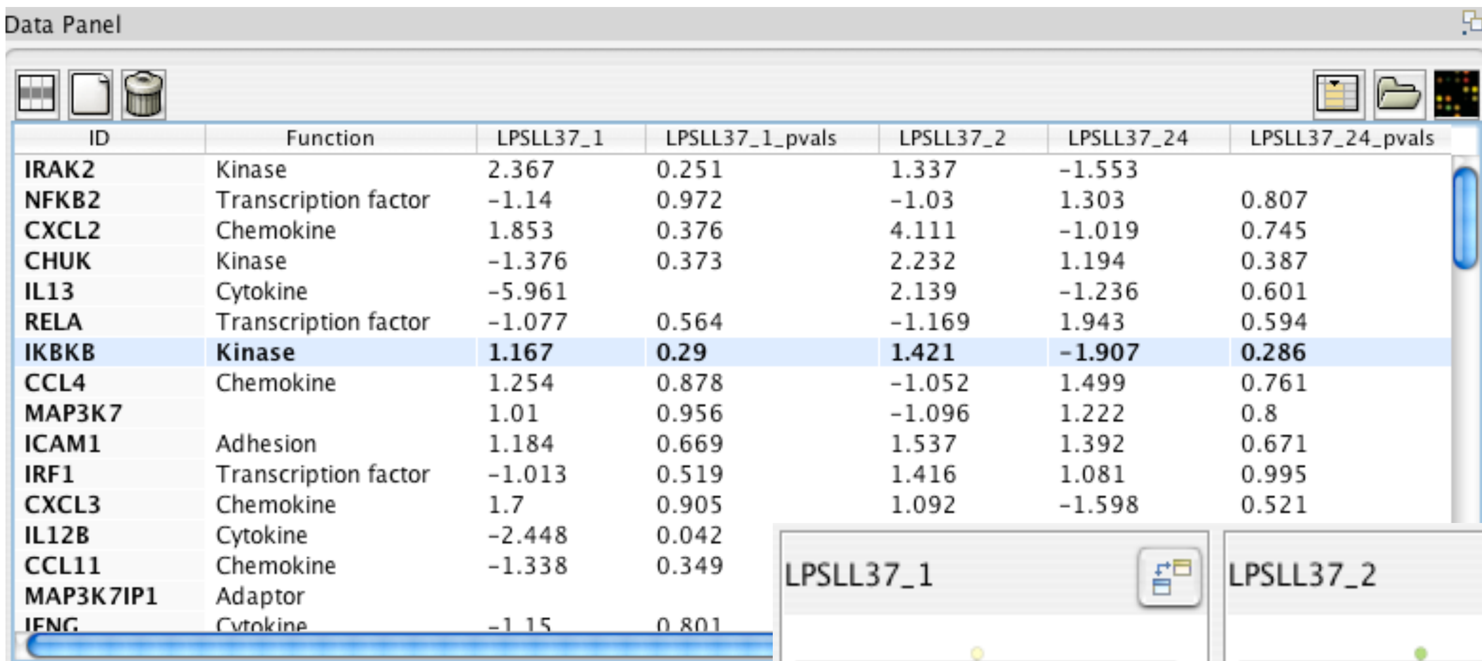
Data Panel

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
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					
IFNG	Cytokine	-1.15	0.801	1.075	1.053	0.521

External representation: replace cognition with perception

Visualization

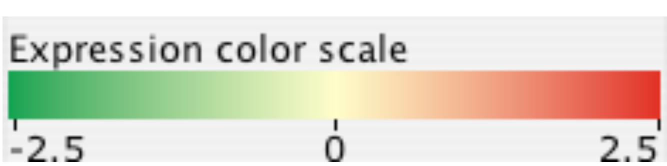
The role of visualization systems is to provide visual representations of datasets that help people **carry out tasks more effectively**.

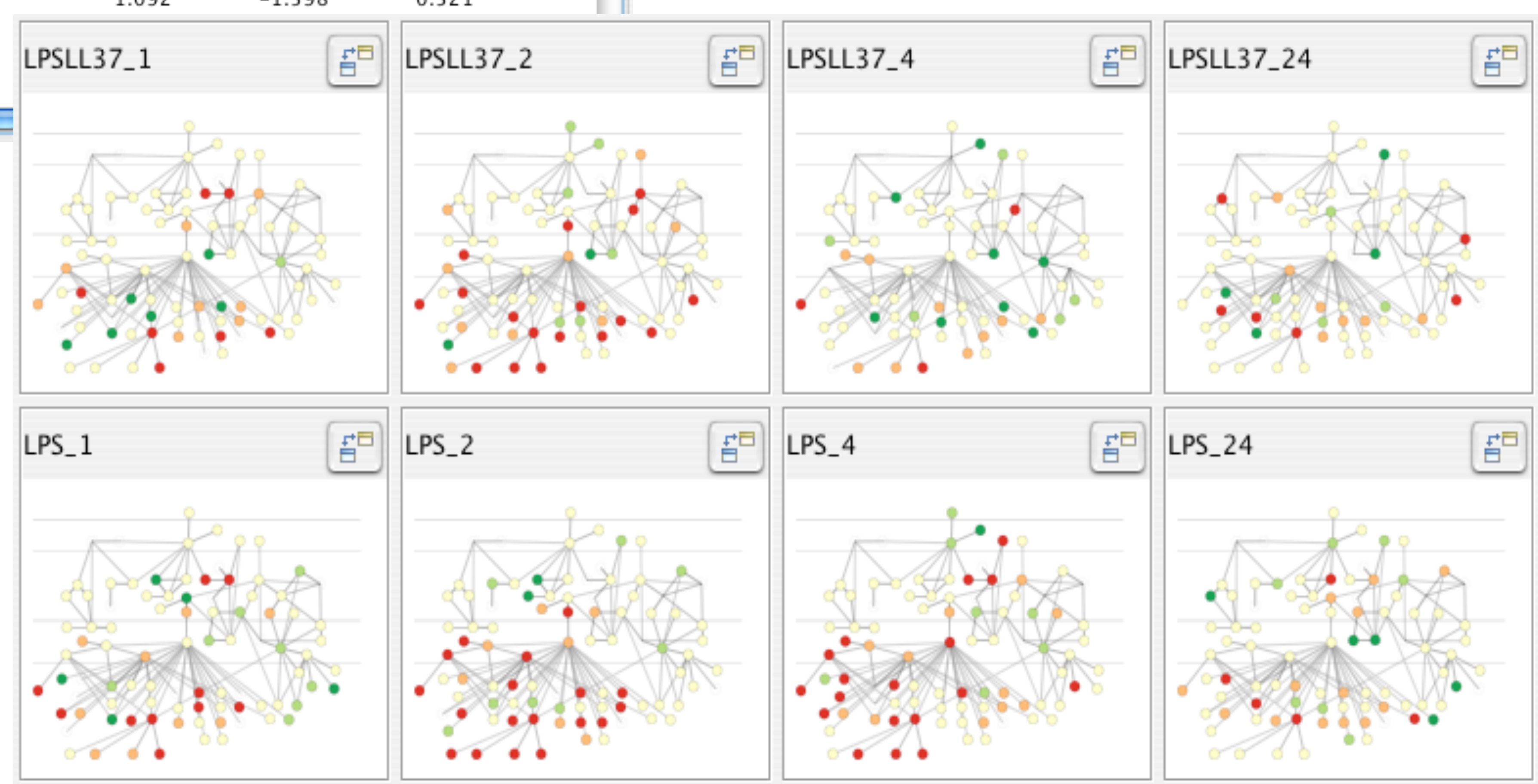


ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
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			
CCL11	Chemokine	-1.338	0.349			
MAP3K7IP1	Adaptor					
IFNG	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.]





What are we visualising?

Major data types & classifications of them

Why are we visualising it?

What is the need for this visualization?

Why do the users need this, and what do they need to be able to do with it?

How can we visualise?

How can we visualize?

The components of a visualization.

Good and bad practices.

What are we visualising?

Major data types & classifications of them

Why are we visualising it?

What is the need for this visualization?

Why do the users need this, and what do they need to be able to do with it?

How can we visualise?

How can we visualize?

The components of a visualization.

Good and bad practices.

What are we visualising?

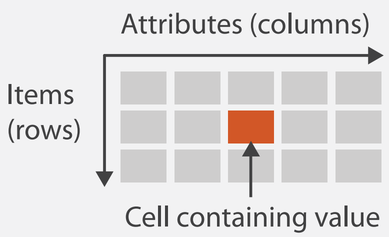
DATA TYPES

STATIC

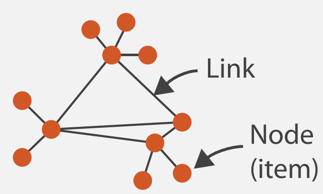
DYNAMIC

DATASET TYPES

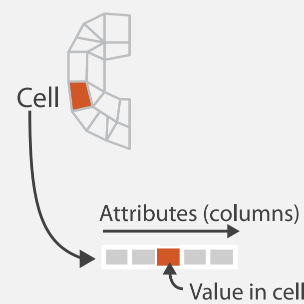
TABLES



NETWORKS



FIELDS (CONTINUOUS)

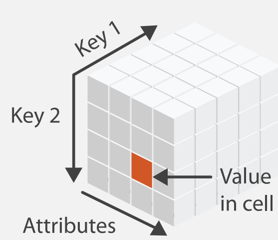


GEOMETRY (SPATIAL)

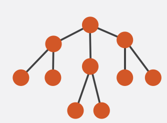
TEXT

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

→ Multidimensional Table



→ Trees

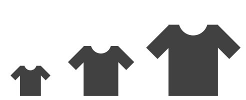


ATTRIBUTE TYPES

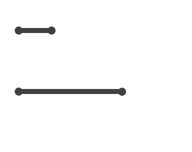
CATEGORICAL

ORDERED

→ Ordinal



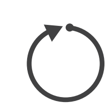
→ Quantitative



→ Sequential

→ Diverging

→ Cyclic



What are we visualising?

STATIC

For static data, we have fixed scales.

We know our data range, therefore scales will not change.

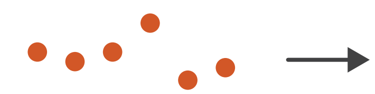
What are we visualising?

STATIC

For static data, we have fixed scales.

We know our data range, therefore scales will not change.

DYNAMIC



For dynamic data, the observed min and max values can change, therefore scales will change.

This can have big consequences for the readability of our visualization.

What are we visualising?

Major data types & classifications of them

Why are we visualising it?

What is the need for this visualization?

Why do the users need this, and what do they need to be able to do with it?

How can we visualise?

How can we visualize?

The components of a visualization.

Good and bad practices.

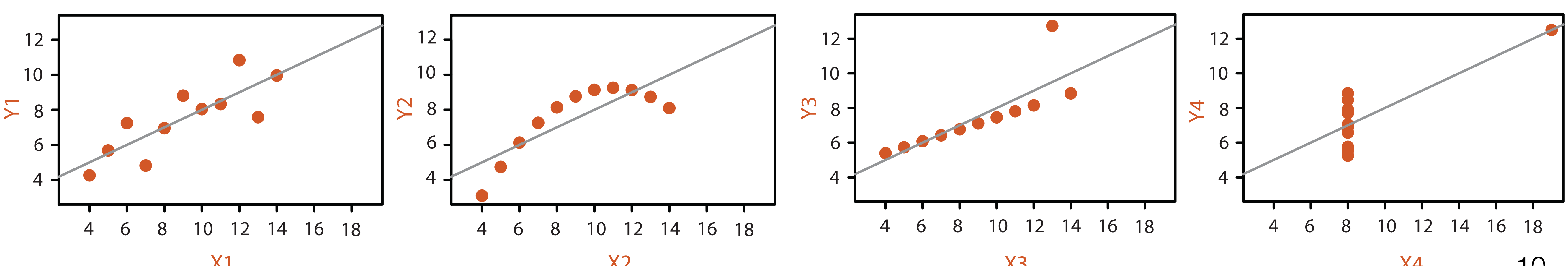
Why are we visualising?

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

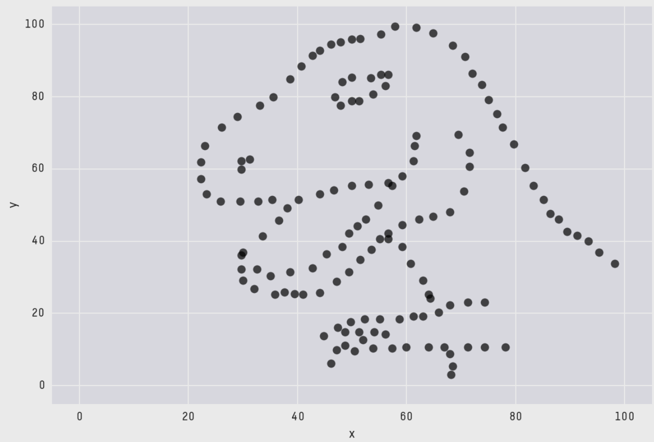
Anscombe's Quartet: Raw Data

	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	

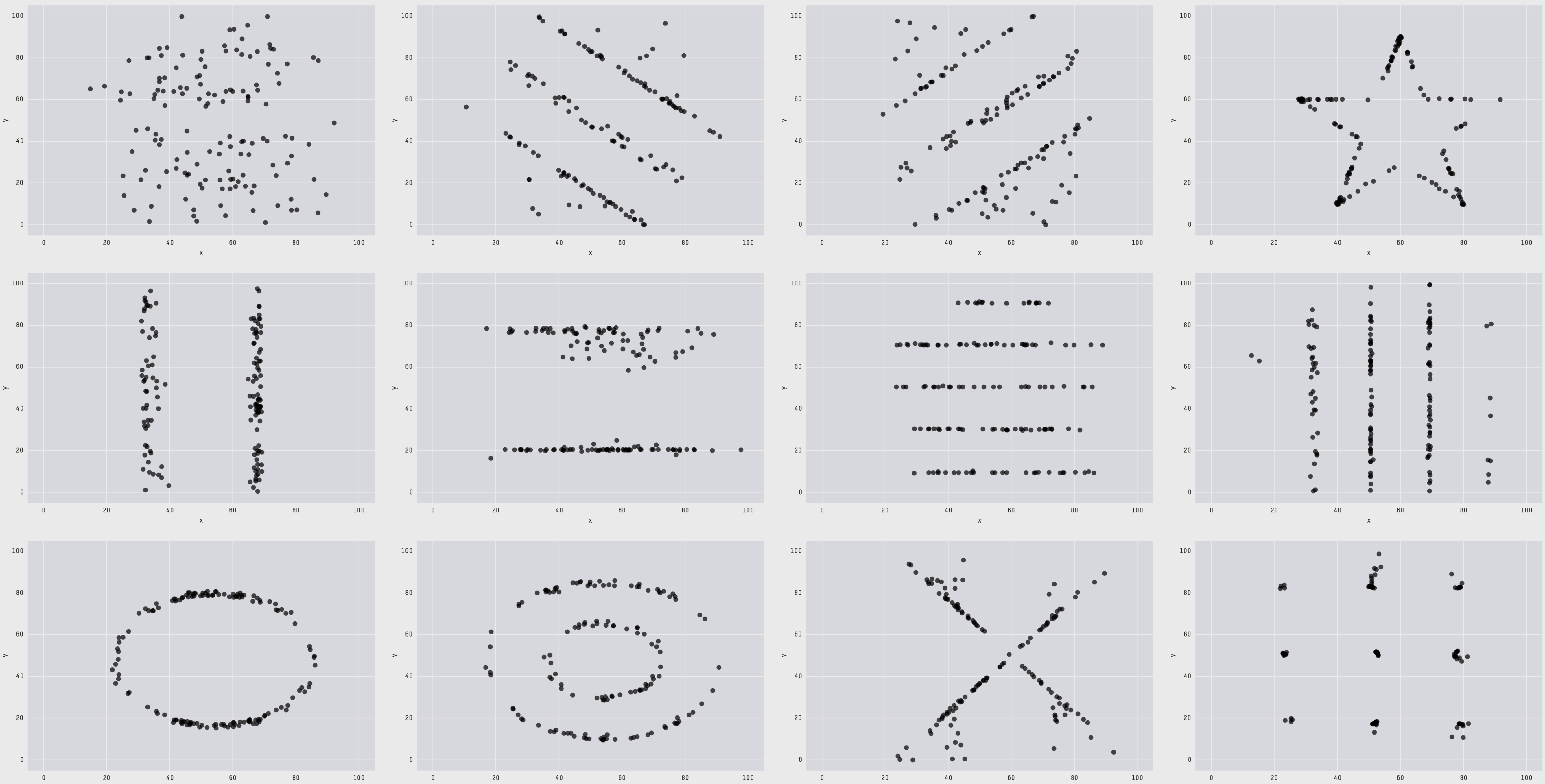
The statistics would lead us to believing that everything is the same



Datasaurus!



X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06



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

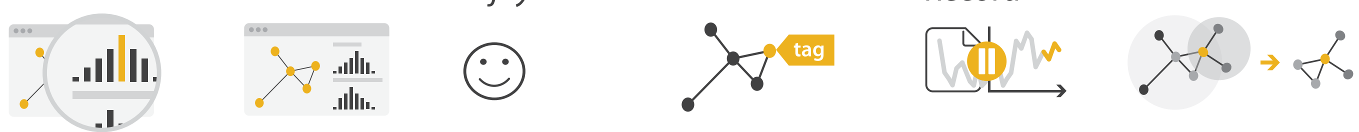
Why are we visualising?

Every visualisation should be thought of as a product of what actions the user needs to take to get to their objective (target)

Actions

Use

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

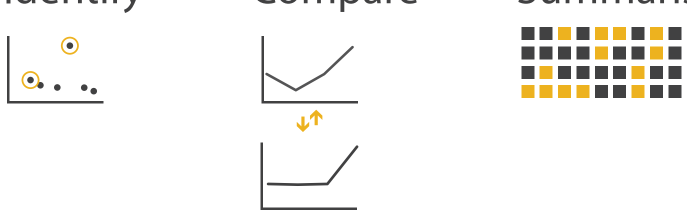


Search

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

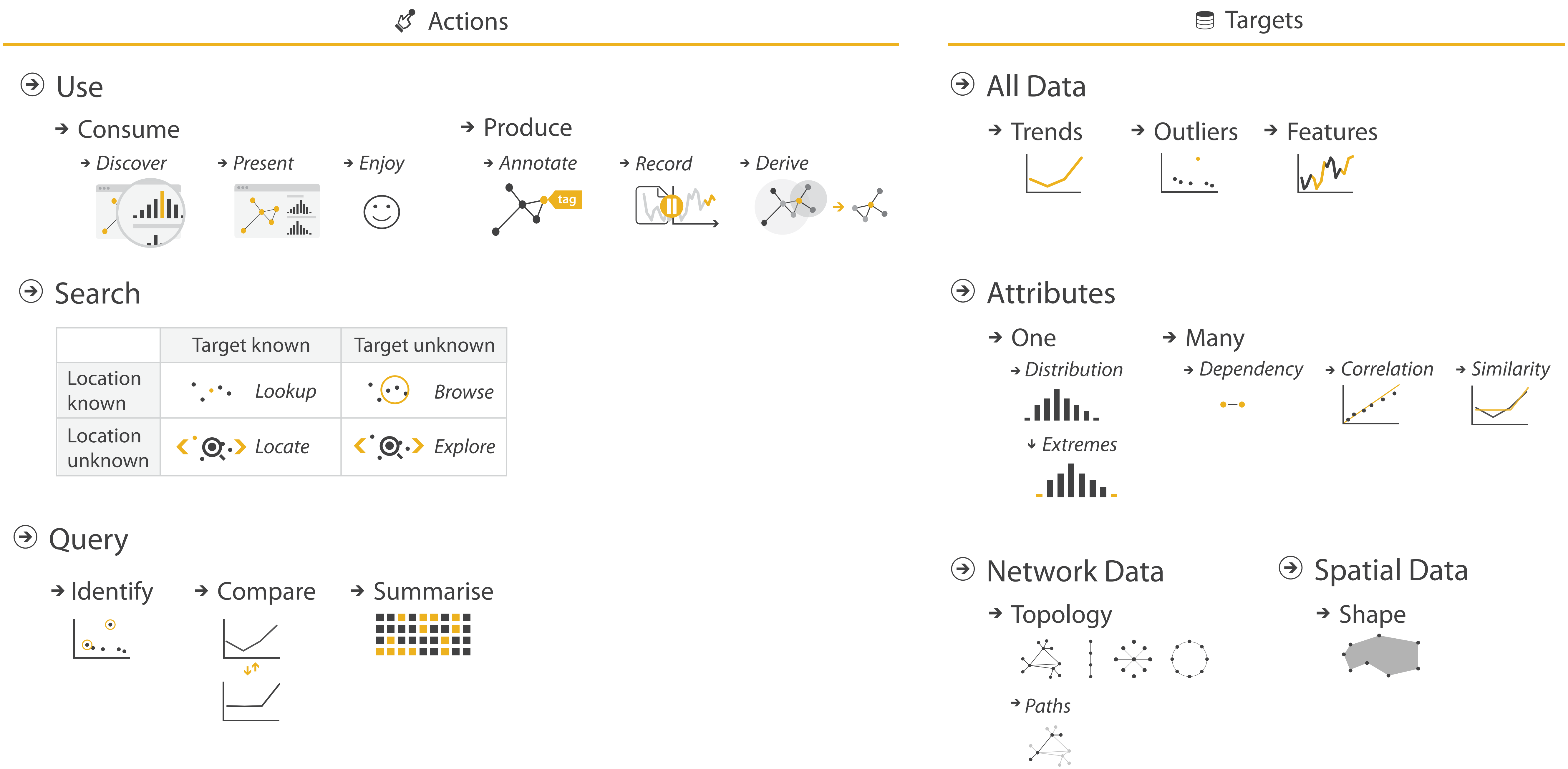
Query

- Identify
- Compare
- Summarise



Why are we visualising?

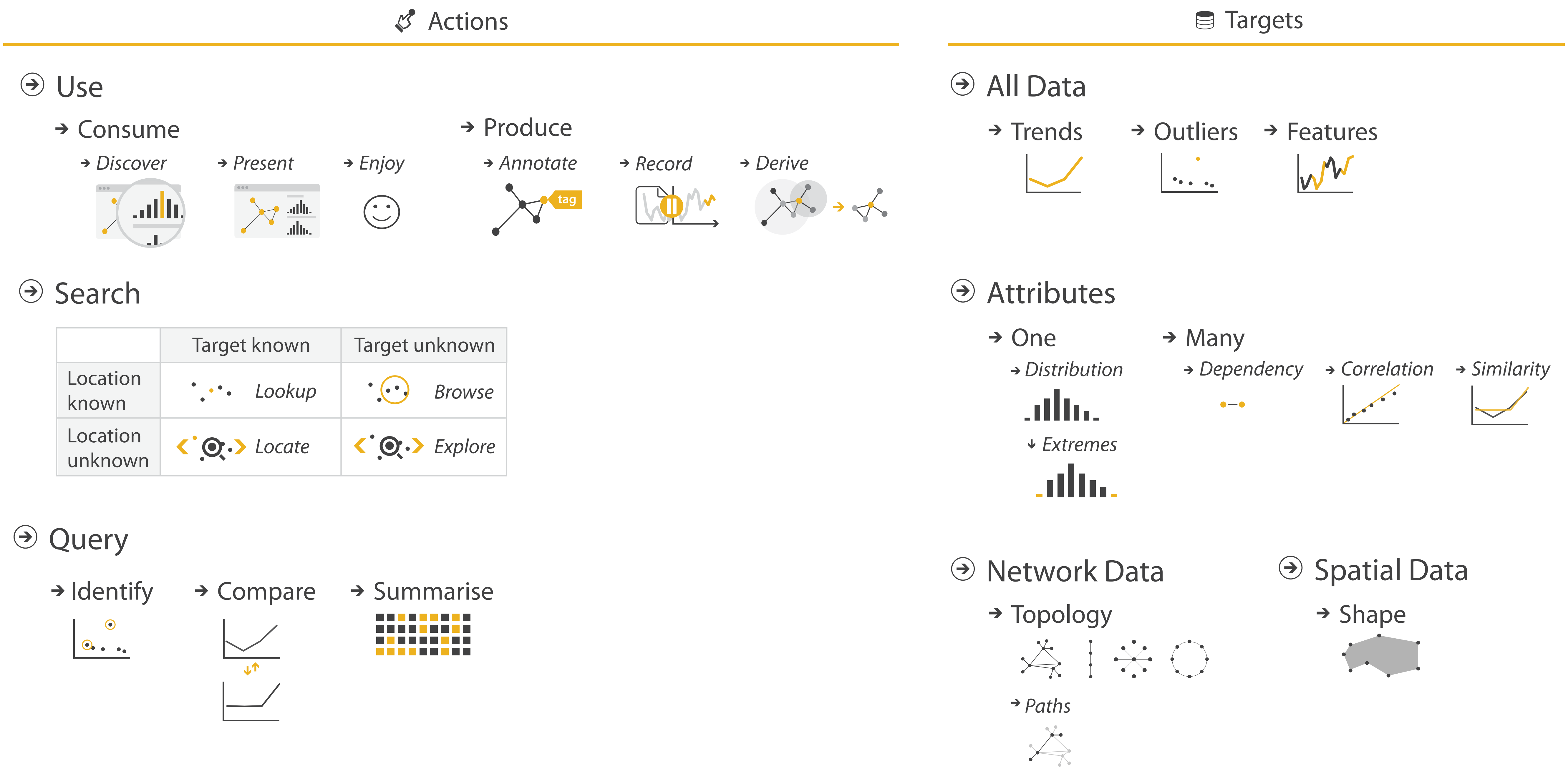
Every visualisation should be thought of as a product of what actions the user needs to take to get to their objective (target)



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

Why are we visualising?

Every visualisation should be thought of as a product of what actions the user needs to take to get to their objective (target)



Always keep in mind why you're doing something. If what you create does not show what you intended, confuses, or misleads, it's time to rethink :)

Why are we visualising?

Given a large matrix, or even a large series of numbers, it’s difficult for humans to ‘see’ patterns in the data.

With a visualisation we want to transition a cognitively demanding task to a perceptual (less demanding) one.

Why are we visualising?

Even in this simple example, it is cognitively demanding to read off all the information.

Category	Sub-Category	Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-45.93	-9,300.00	-16,000.00	-7,600.00
	Chairs & Chairmats	42,900.00	41,300.10	41,000.00	25,600.00
	Office Furnishings	12,000.00	27,300.10	42,000.00	18,600.00
	Tables	-12,300.00	-35,400.10	-43,000.00	-8,000.00
Technology	Computer Peripherals	14,100.56	45,300.00	17,000.00	17,300.00
	Copiers & Fax	41,300.00	-28,600.10	29,000.00	68,100.00
	Office Machines	51,400.00	180,300.10	39,000.00	36,500.00
	Comms (Telephones)	49,700.00	120,400.10	86,000.00	-59,800.00

What is the goal of this representation?

Why are we visualising?

We can improve by using ‘pop-out’ to bring attention to negative values.

Category	Sub-Category	Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-45.93	-9,300.00	-16,000.00	-7,600.00
	Chairs & Chairmats	42,900.00	41,300.10	41,000.00	25,600.00
	Office Furnishings	12,000.00	27,300.10	42,000.00	18,600.00
	Tables	-12,300.00	-35,400.10	-43,000.00	-8,000.00
Technology	Computer Peripherals	14,100.56	45,300.00	17,000.00	17,300.00
	Copiers & Fax	41,300.00	28,600.10	29,000.00	68,100.00
	Office Machines	51,400.00	180,300.10	39,000.00	36,500.00
	Comms (Telephones)	49,700.00	120,400.10	86,000.00	-59,800.00

Why are we visualising?

Or, adding some additional indicators can provide an idea of intensity.

Category	Sub-Category	Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-45.93	-9,300.00	-16,000.00	-7,600.00
	Chairs & Chairmats	42,900.00	41,300.10	41,000.00	25,600.00
	Office Furnishings	12,000.00	27,300.10	42,000.00	18,600.00
	Tables	-12,300.00	-35,400.10	-43,000.00	-8,000.00
Technology	Computer Peripherals	14,100.56	45,300.00	17,000.00	17,300.00
	Copiers & Fax	41,300.00	28,600.10	29,000.00	68,100.00
	Office Machines	51,400.00	180,300.10	39,000.00	36,500.00
	Comms (Telephones)	49,700.00	120,400.10	86,000.00	-59,800.00

Why are we visualising?

Or, adding some additional indicators can provide an idea of intensity.

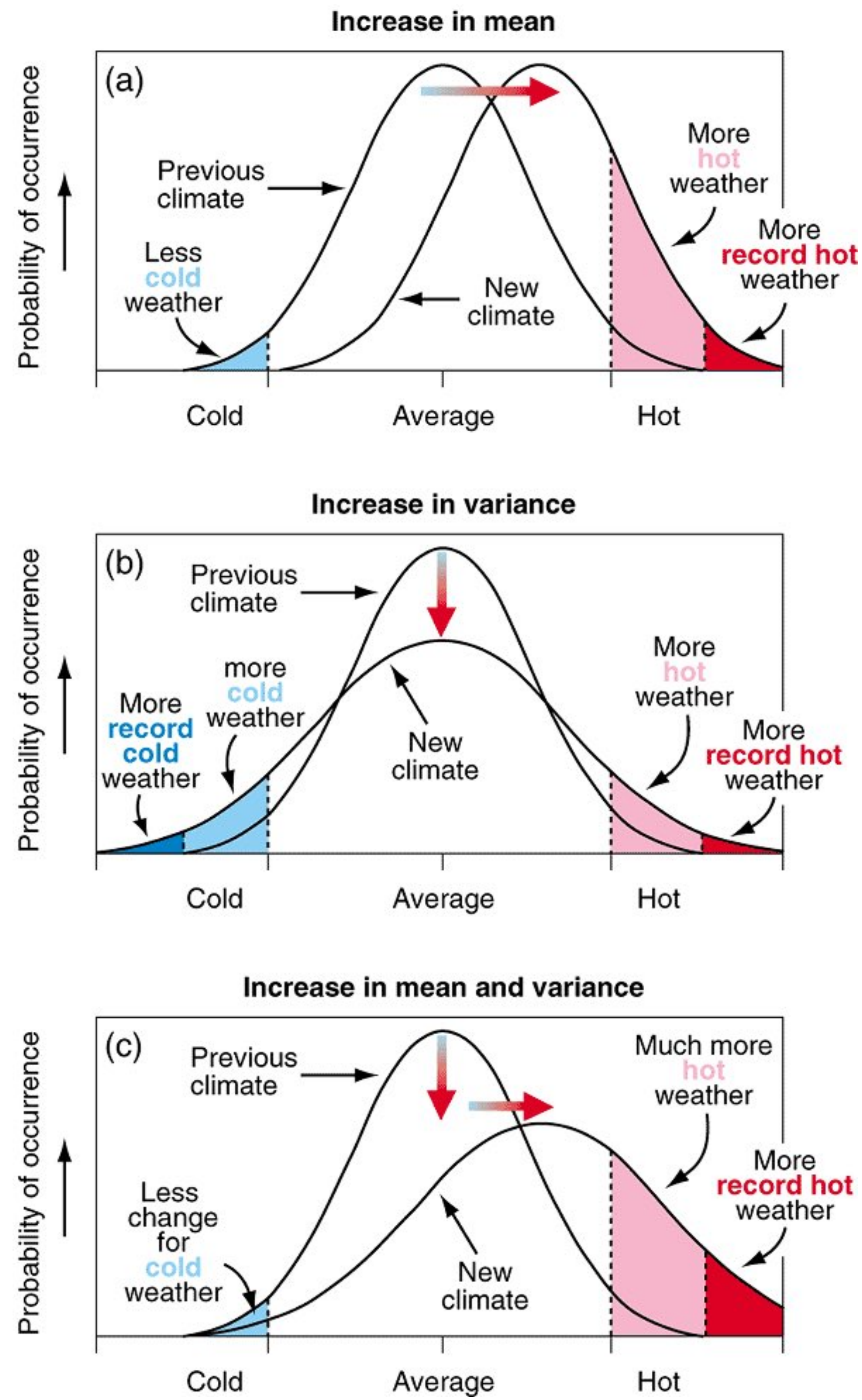
Category	Sub-Category	Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-45.93	-9,300.00	-16,000.00	-7,600.00
	Chairs & Chairmats	42,900.00	41,300.10	41,000.00	25,600.00
	Office Furnishings	12,000.00	27,300.10	42,000.00	18,600.00
	Tables	-12,300.00	-35,400.10	-43,000.00	-8,000.00
Technology	Computer Peripherals	14,100.56	45,300.00	17,000.00	17,300.00
	Copiers & Fax	41,300.00	28,600.10	29,000.00	68,100.00
	Office Machines	51,400.00	180,300.10	39,000.00	36,500.00
	Comms (Telephones)	49,700.00	120,400.10	86,000.00	-59,800.00

How we present information depends on why we are presenting it...

Why are we visualising?

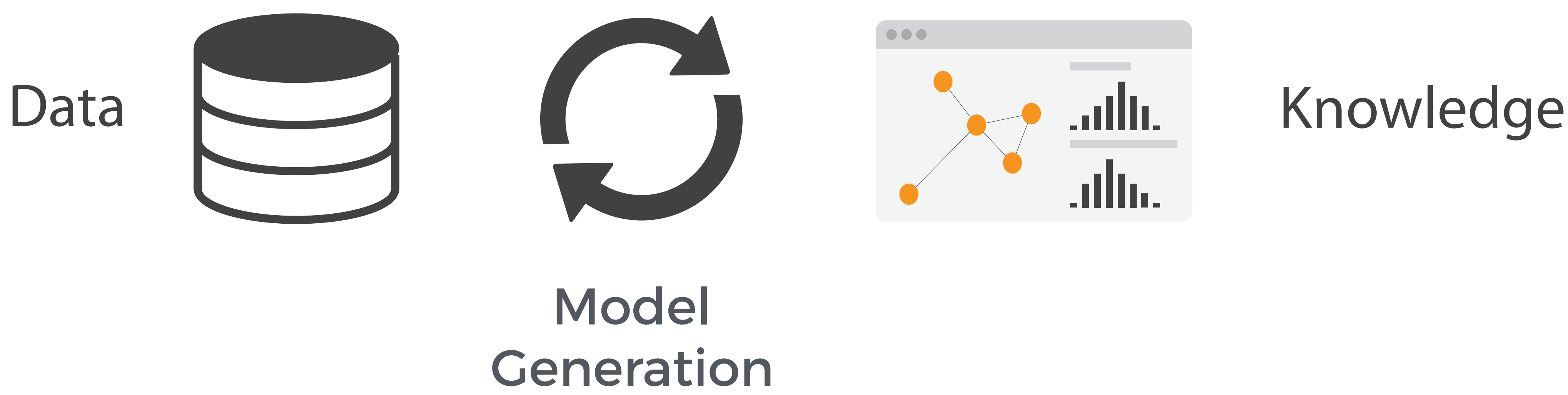
Sometimes it is to communicate information

We can use visualisation to better communicate concepts that aren't easily explained using text alone.



https://www.ipcc.ch/ipccreports/tar/wg1/fig2-32.htm

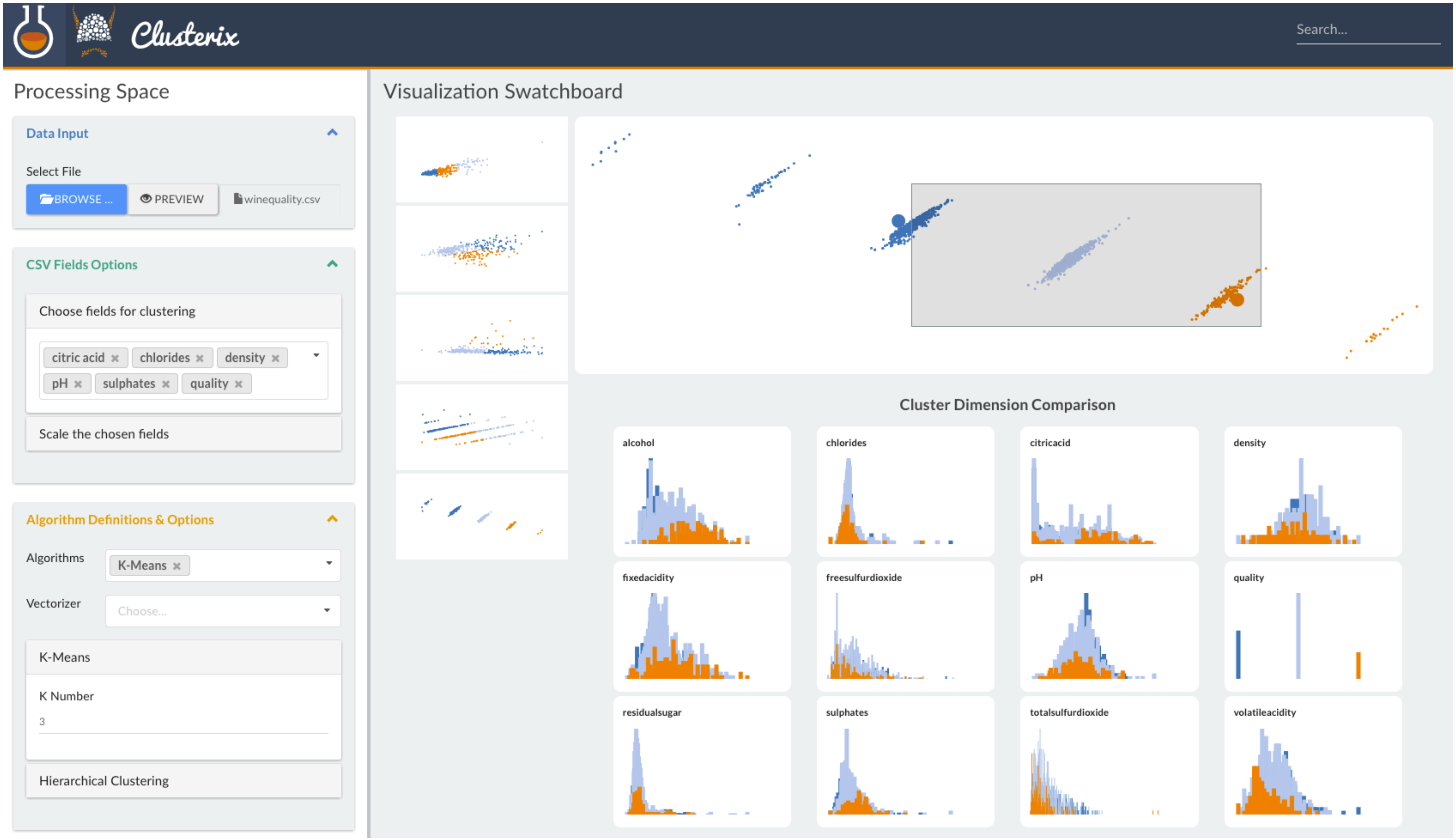
Sometimes visualization is needed to figure out what the best way to represent a data set can be...combining **analytics**, **visualization**, and **human reasoning**.



This is **visual analytics**.

Why are we visualising?

Discovery and Exploration



Joint Work with Ilias Koutsakis and Gilles Louppe @ CERN

What are we visualising?

Major data types & classifications of them

Why are we visualising it?

What is the need for this visualization?

Why do the users need this, and what do they need to be able to do with it?

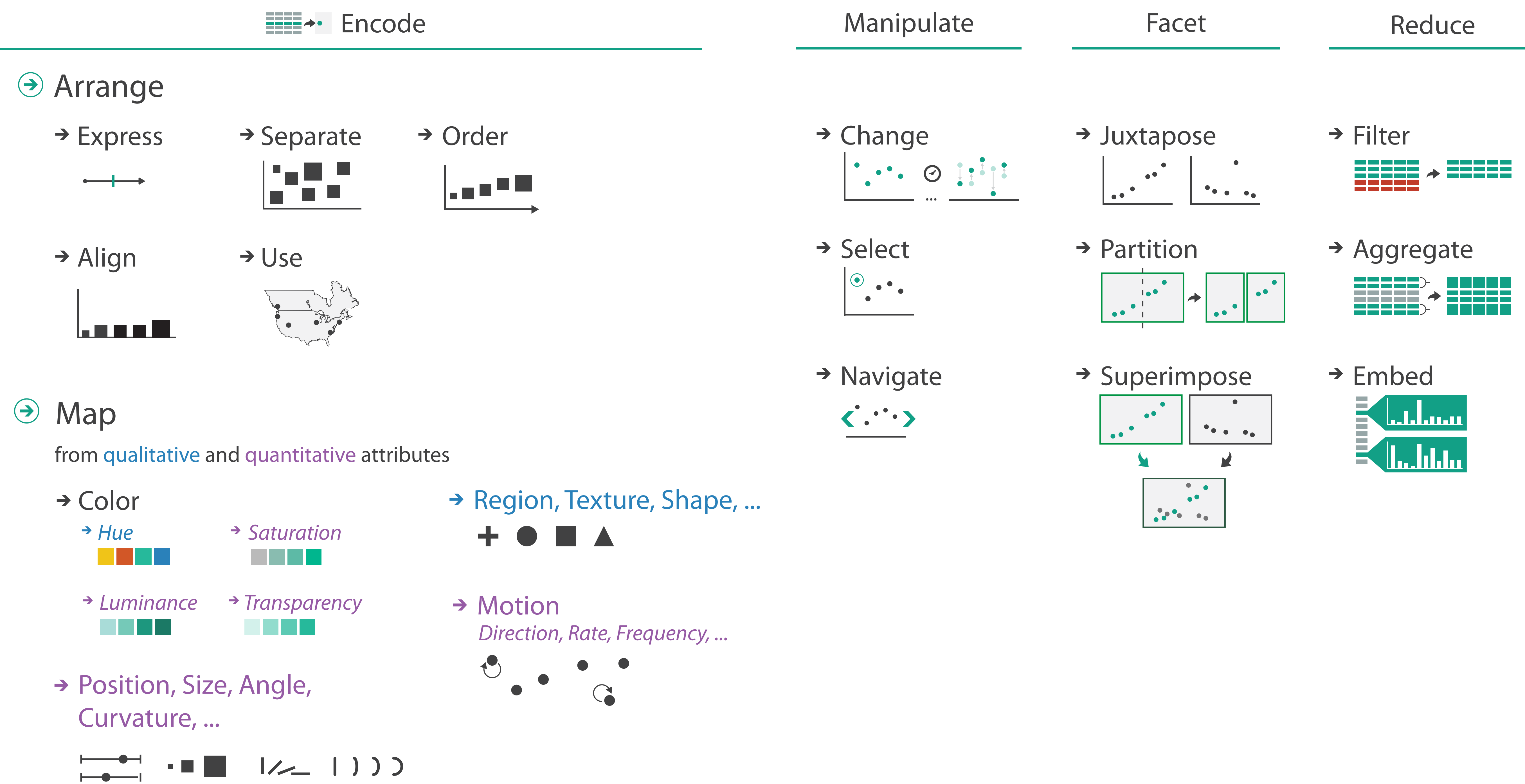
How can we visualise?

How can we visualize?

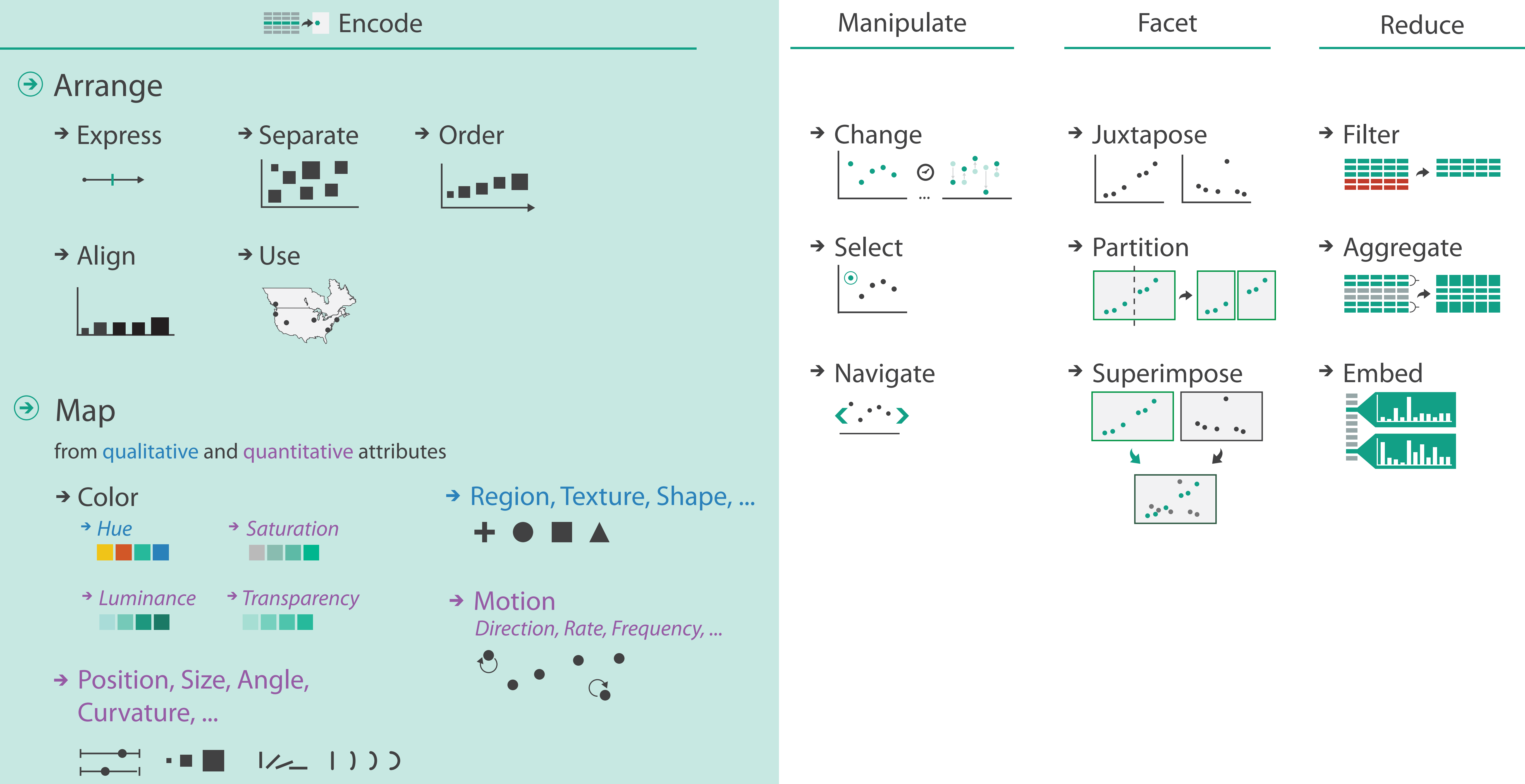
The components of a visualization.

Good and bad practices.

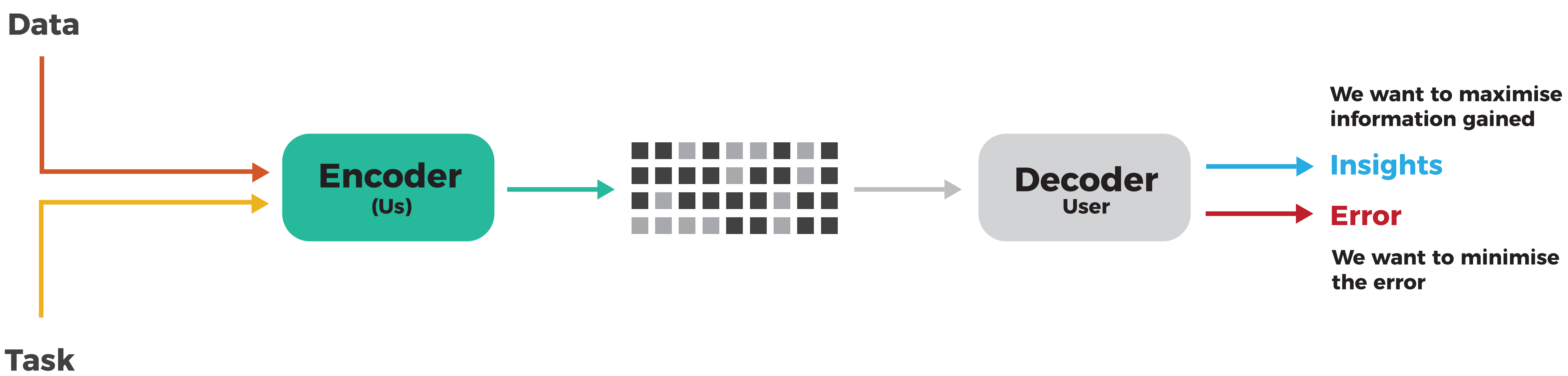
How can you encode information optimally?



How can you encode information optimally?



If we don't follow grammatical rules or spell correctly, the meaning of text can be lost.



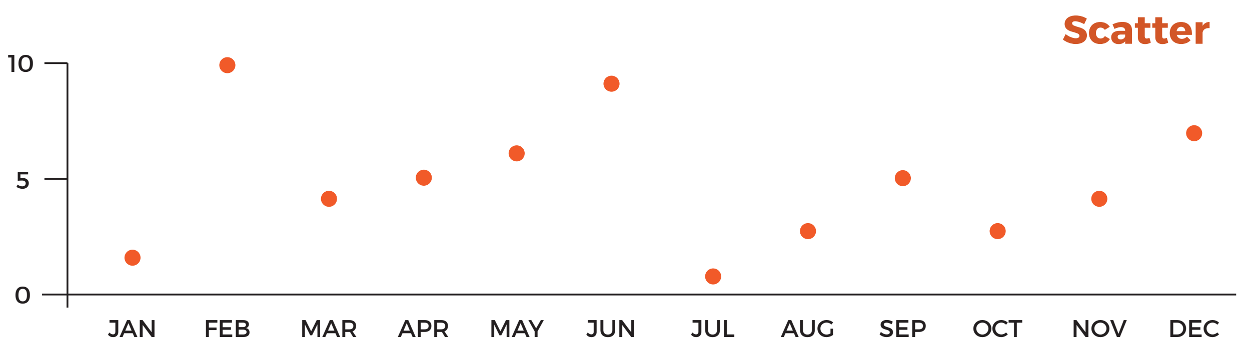
The same applies for visualisations. We can compose visualisations using a vocabulary (shapes, colour, texture,...), and a grammar. If we learn these, we can do better when it comes to communicating visually.

Graphs are like jokes.
If you have to explain them, they didn't work.

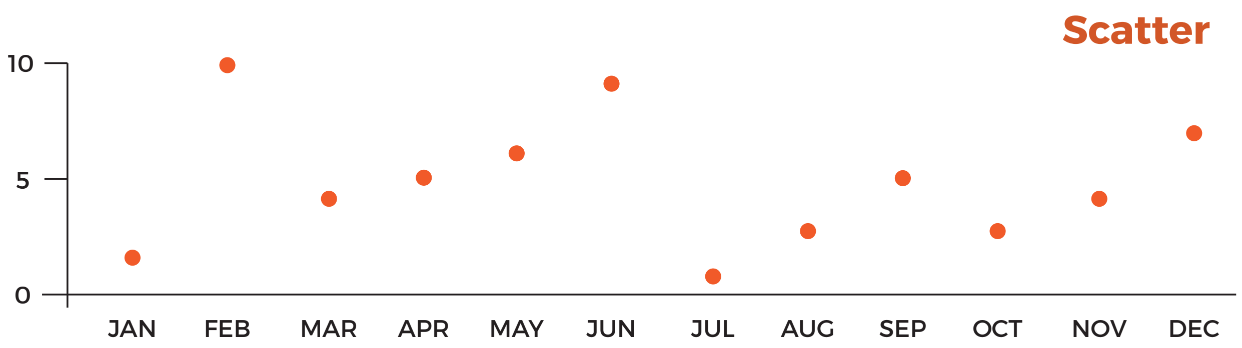
Anon.

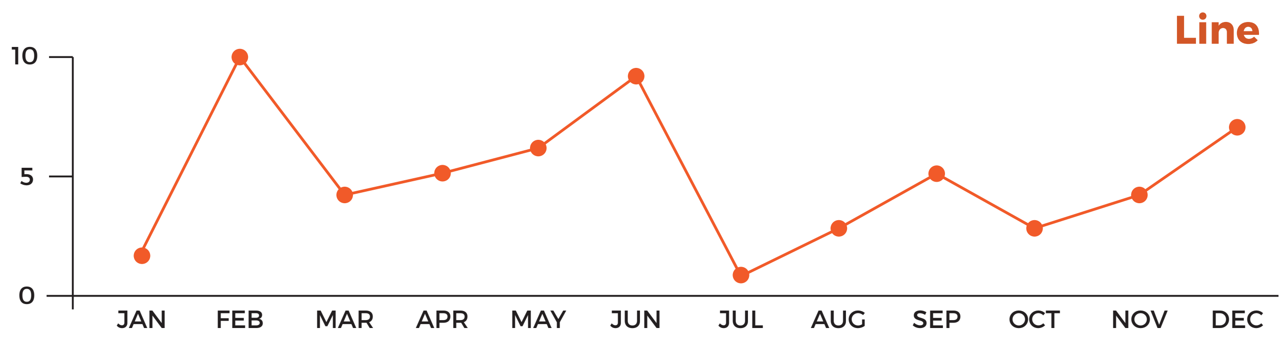
JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

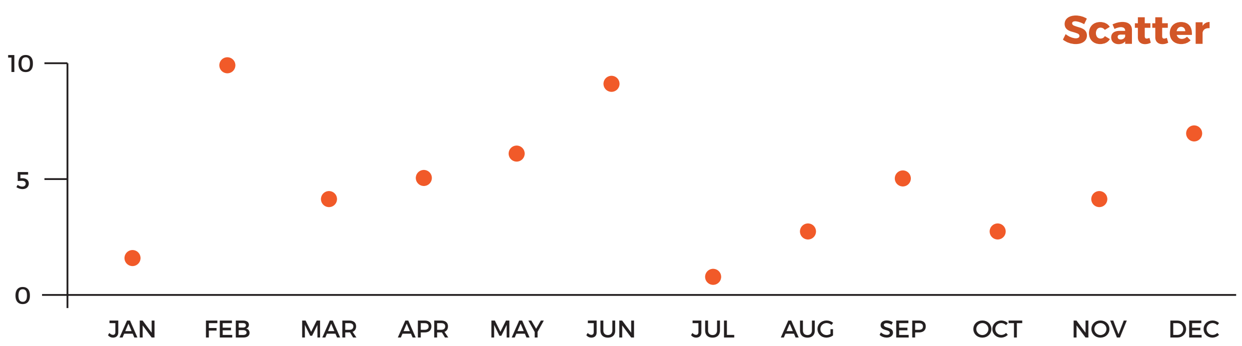


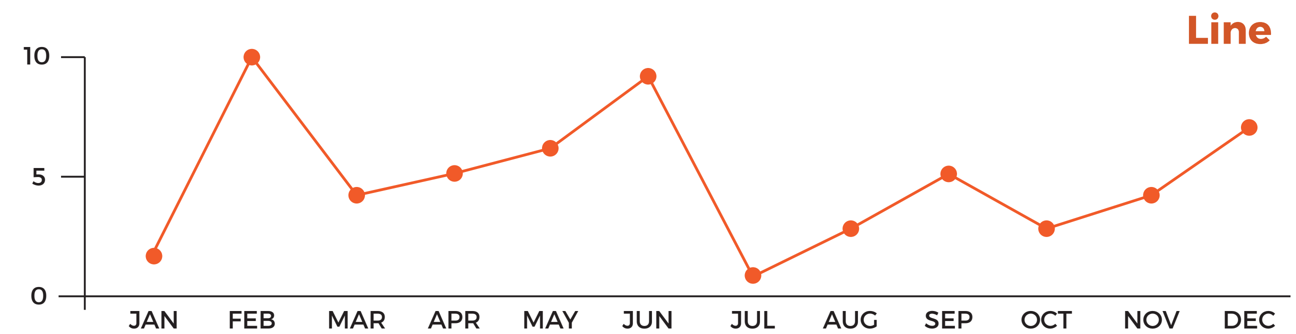
JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

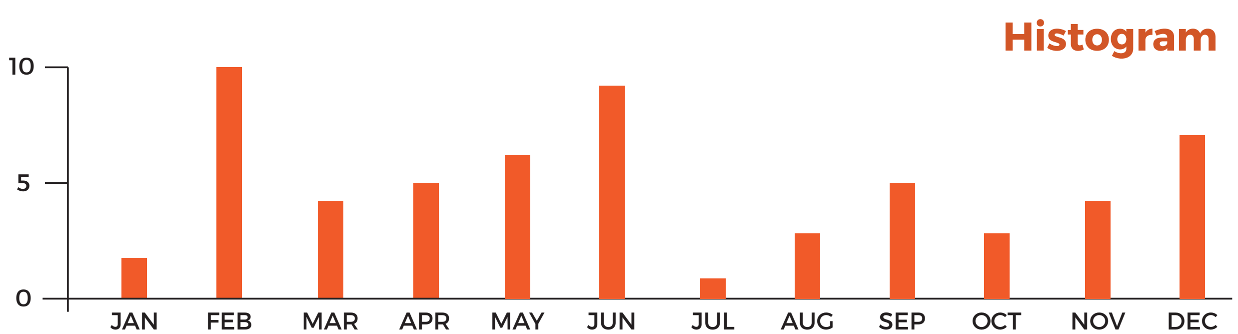




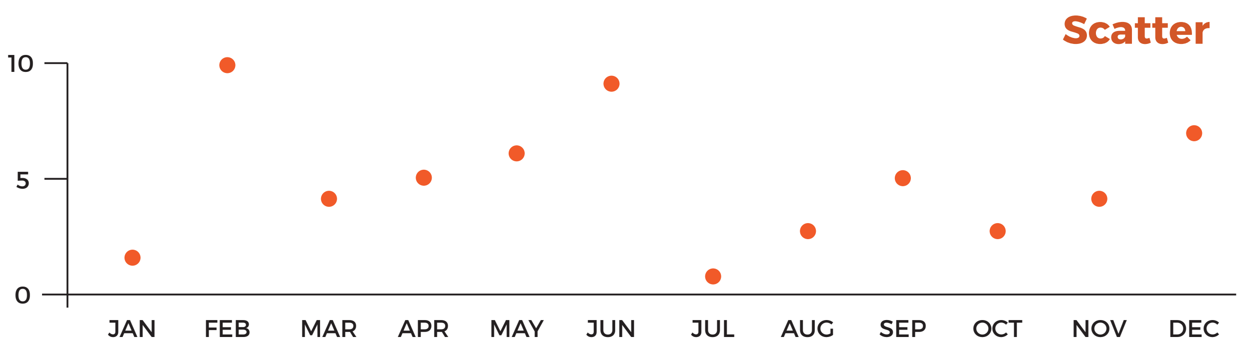
JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

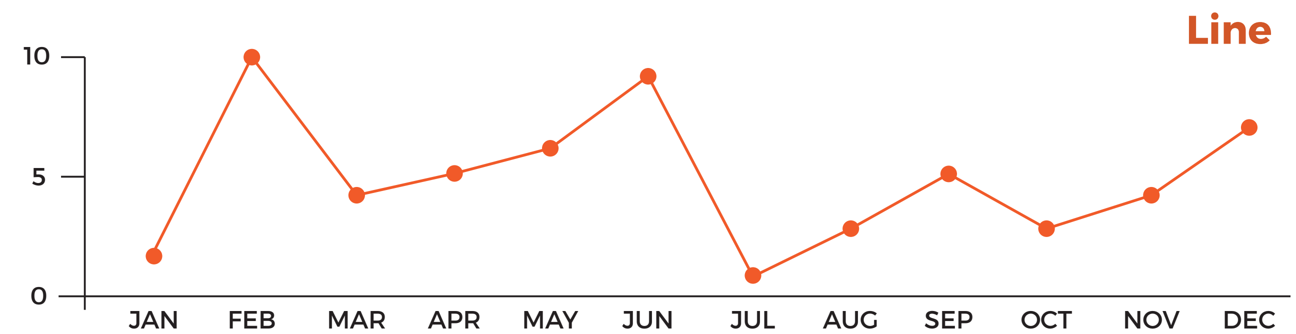


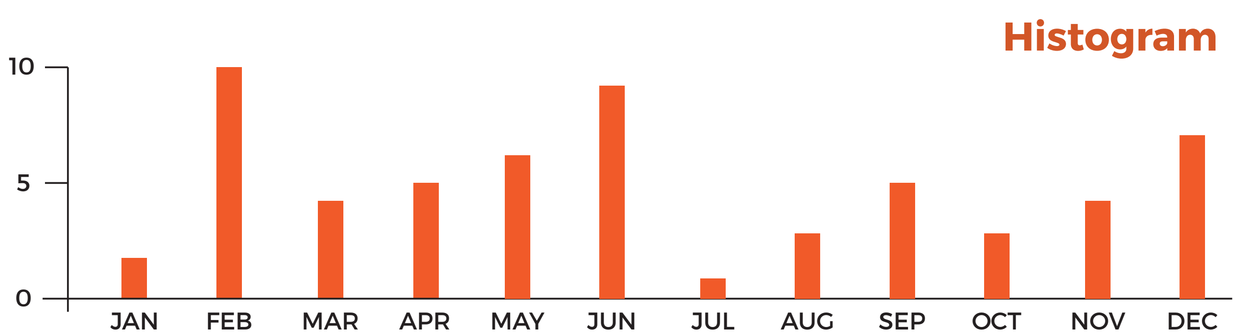


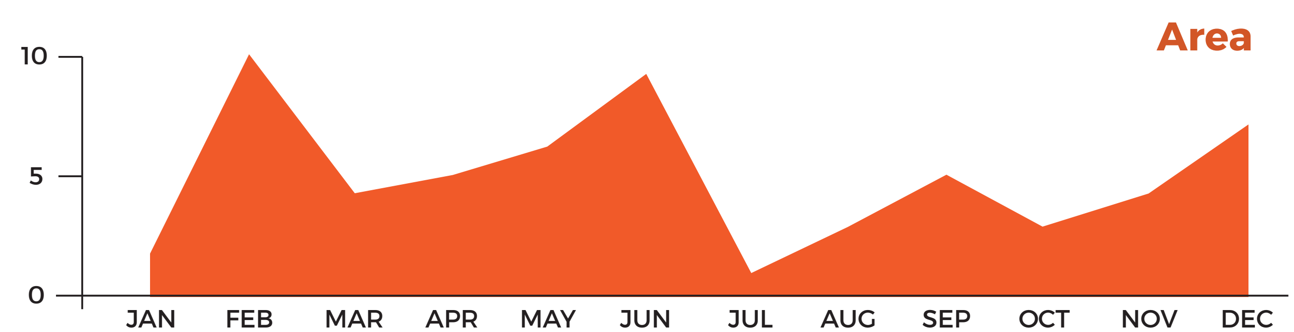


JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7



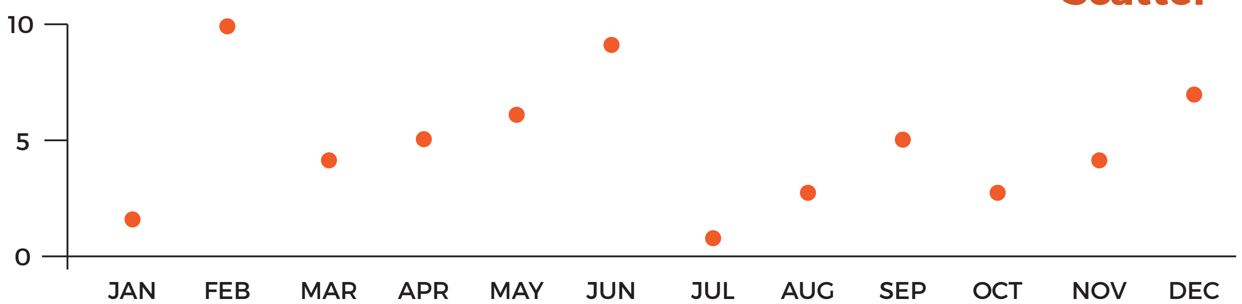




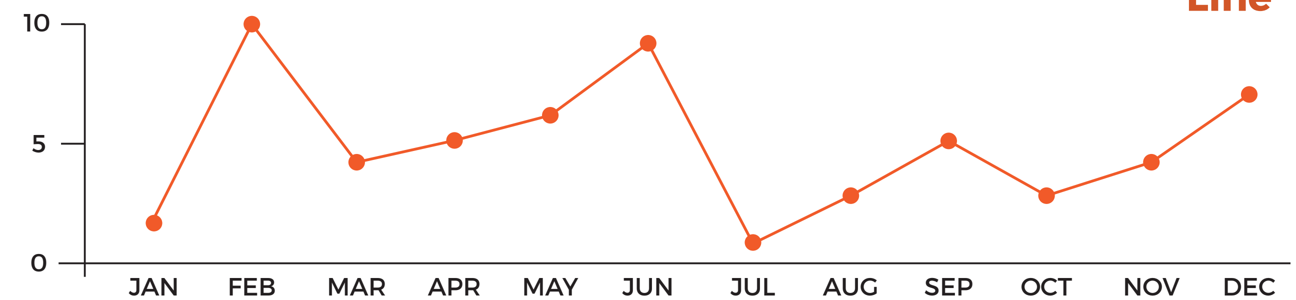


JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

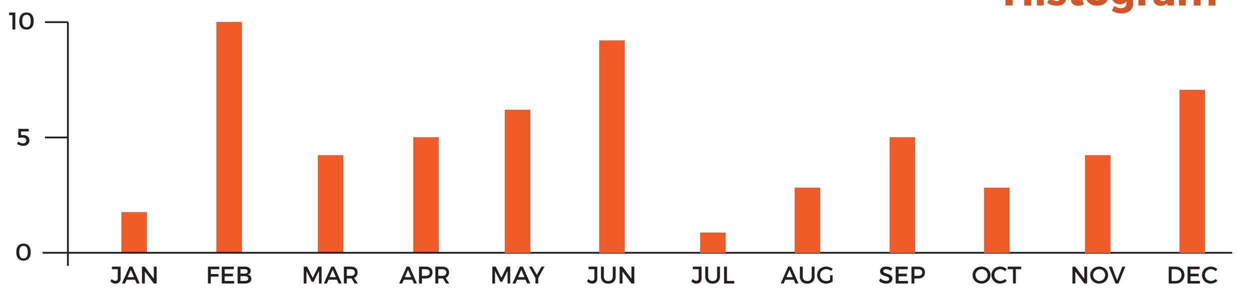
Scatter



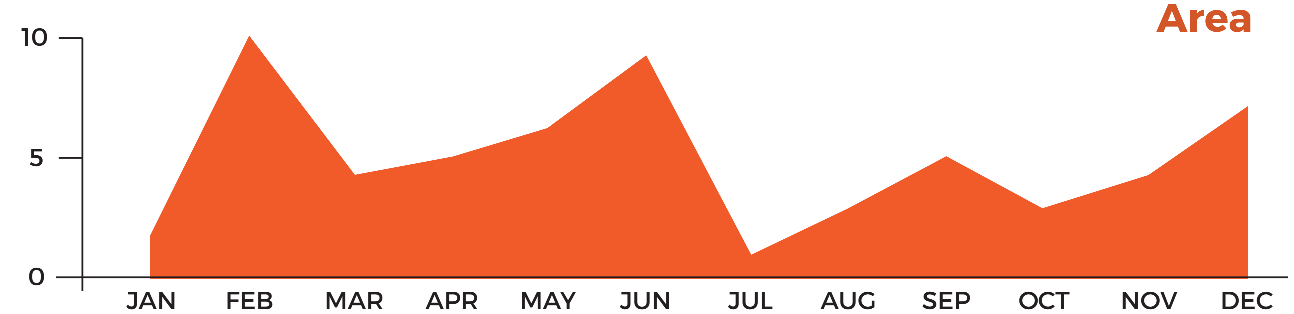
Line



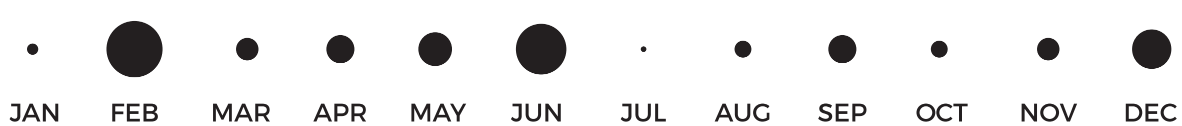
Histogram



Area

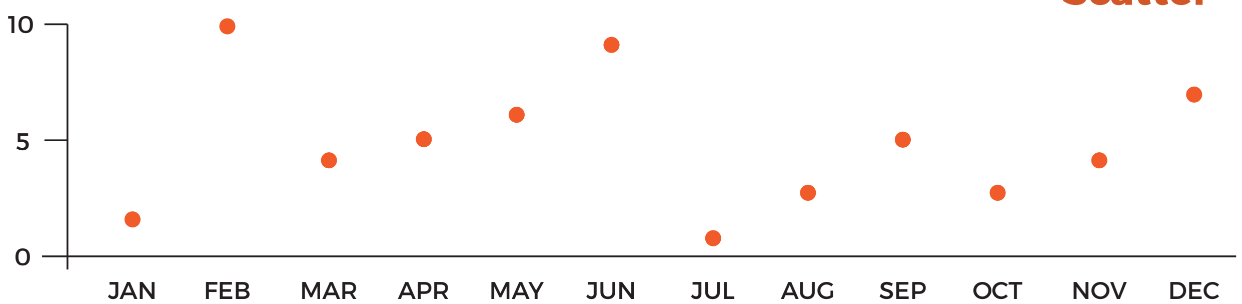


Size

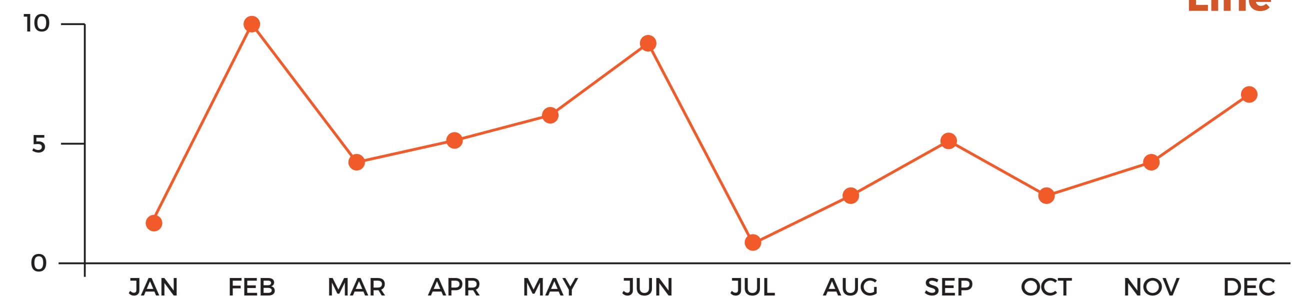


JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

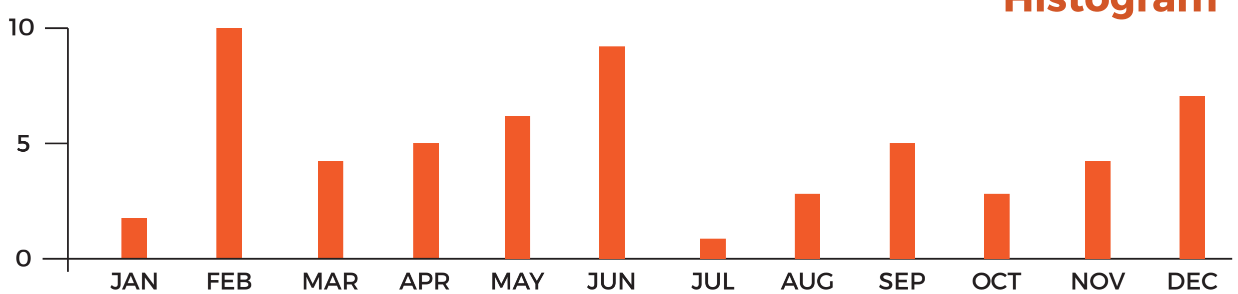
Scatter



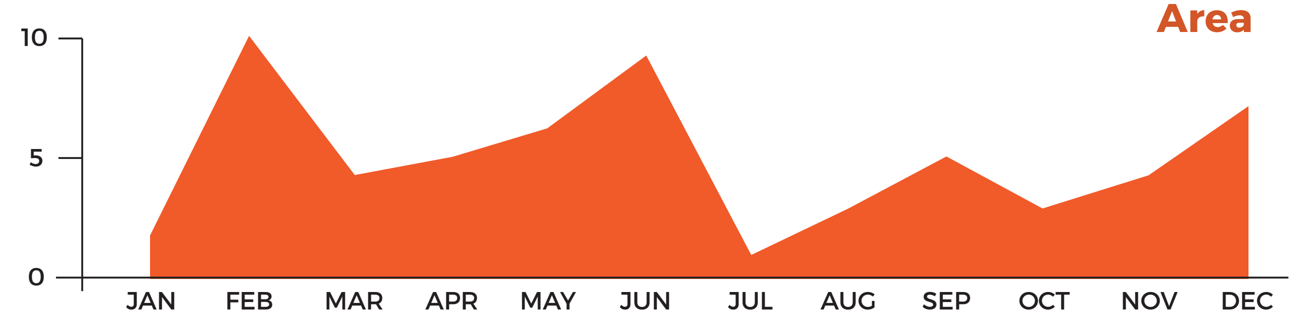
Line



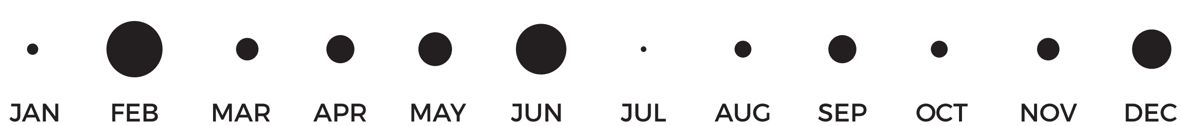
Histogram



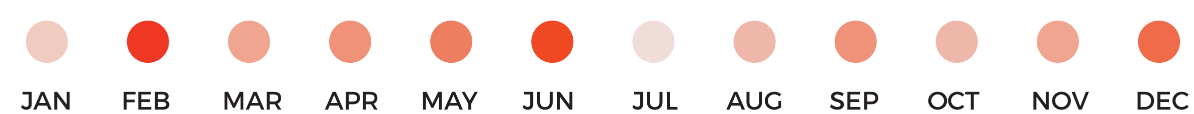
Area



Size

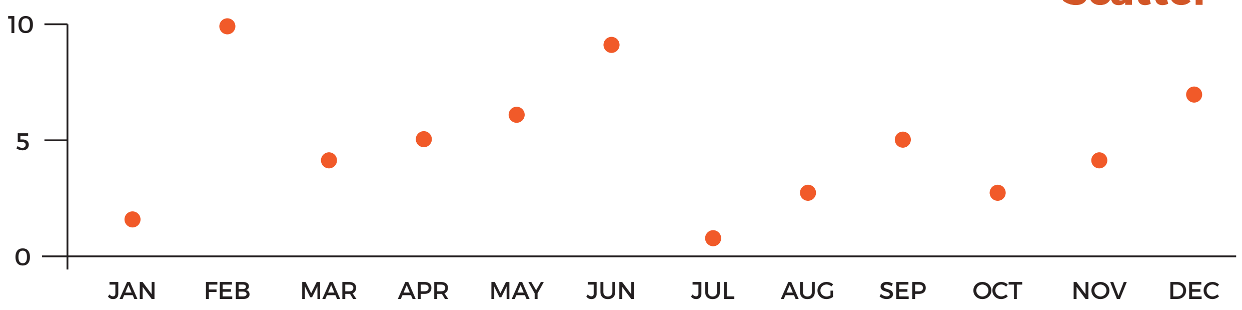


Saturation

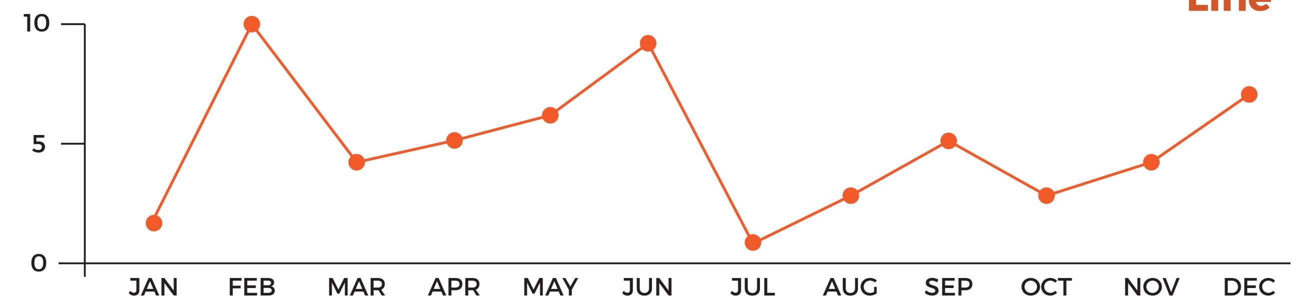


JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

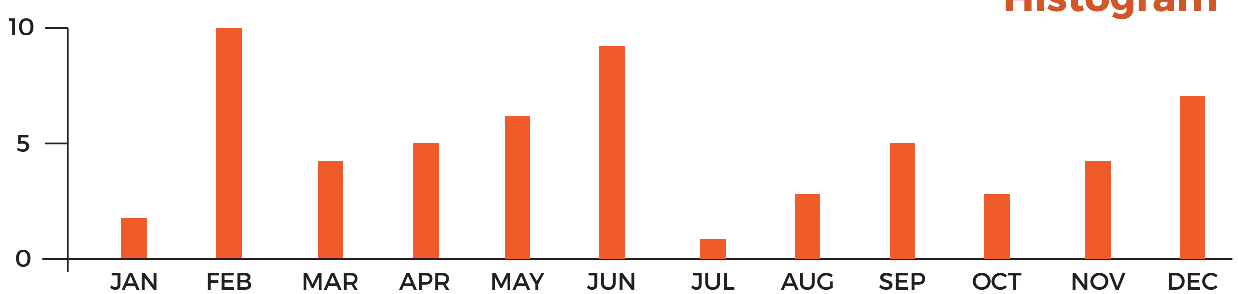
Scatter



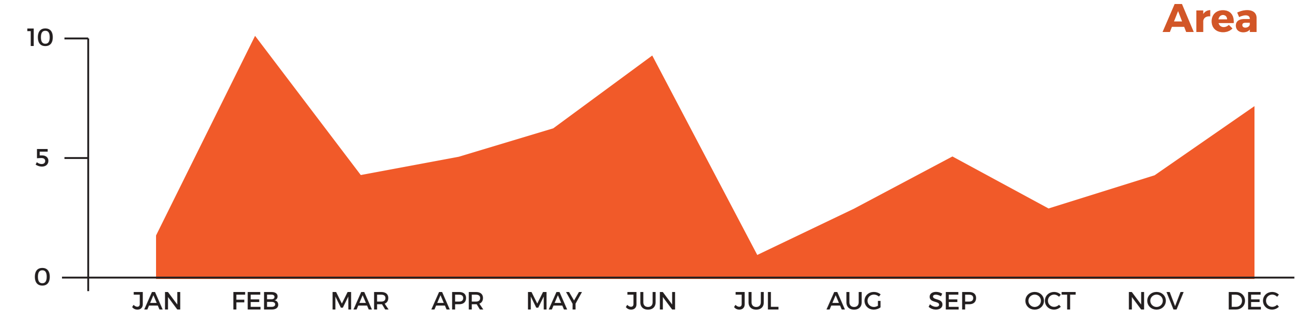
Line



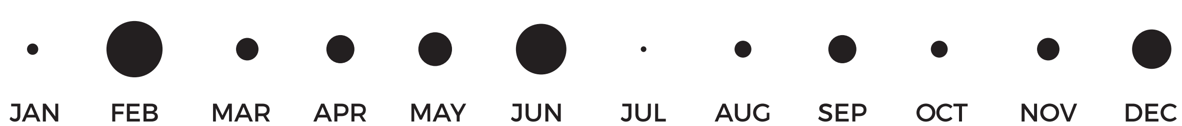
Histogram



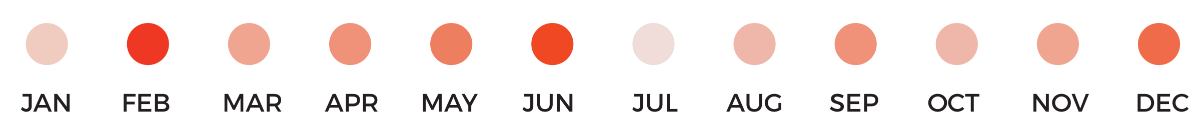
Area



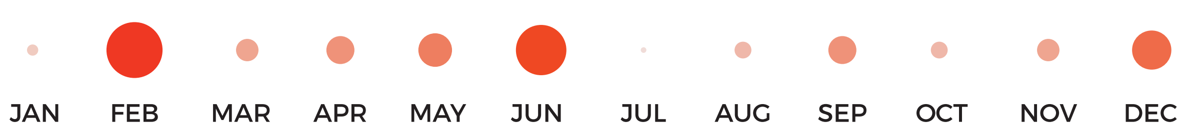
Size



Saturation

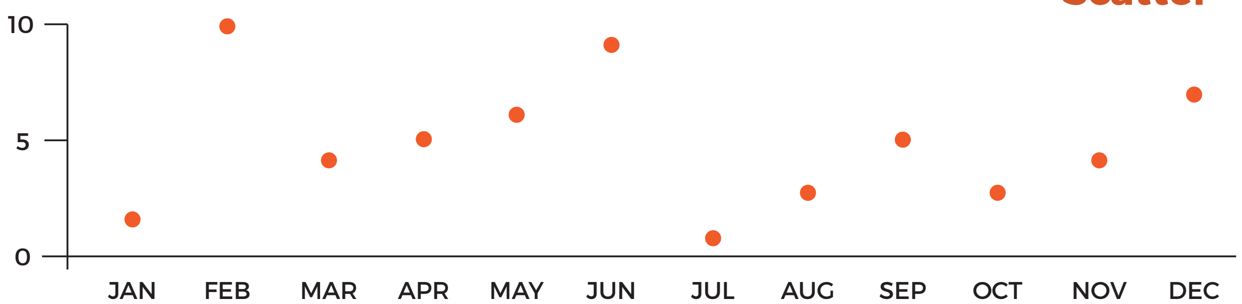


Size & Saturation

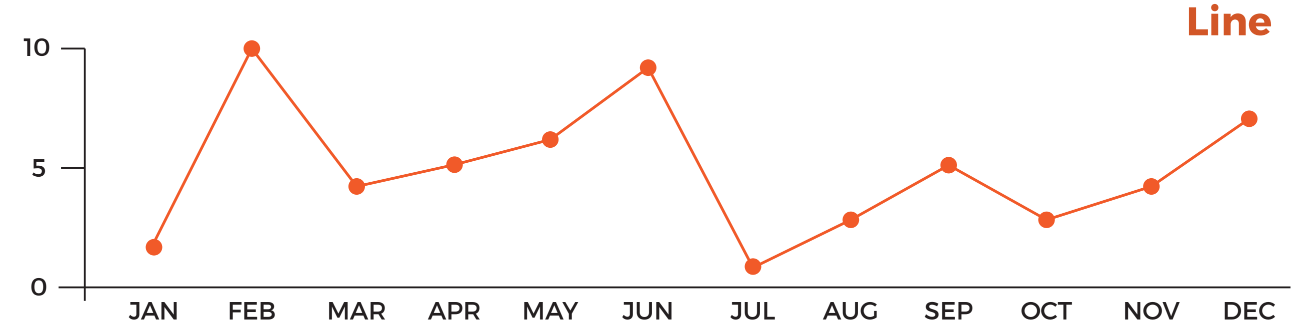


JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

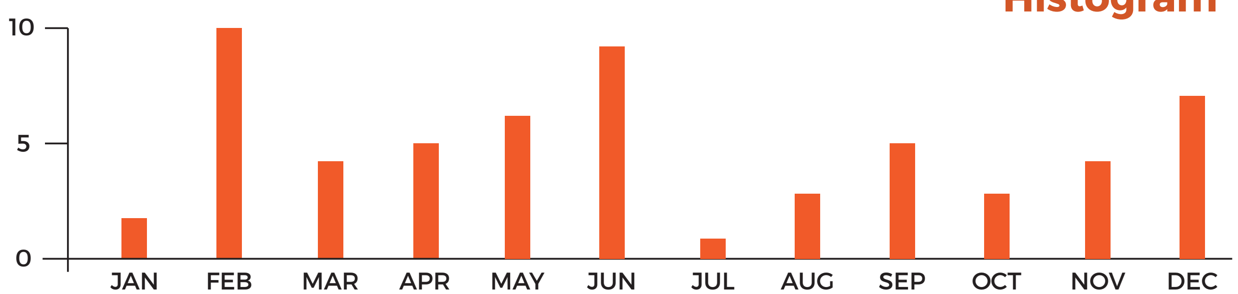
Scatter



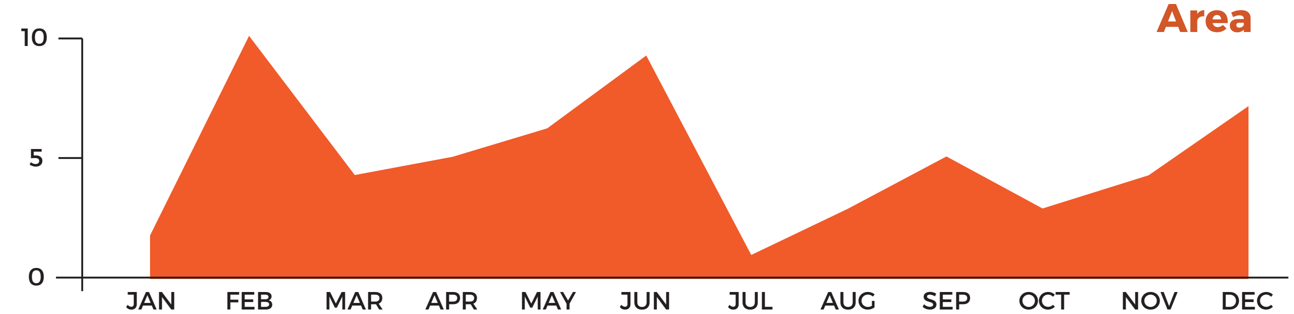
Line



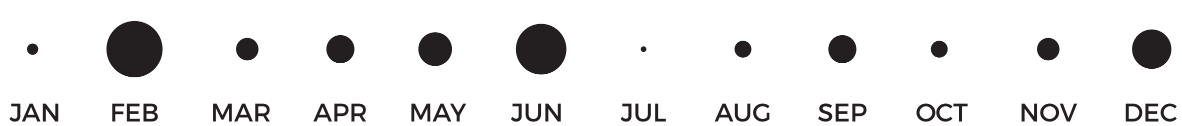
Histogram



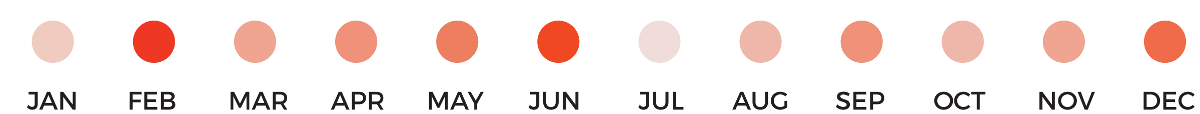
Area



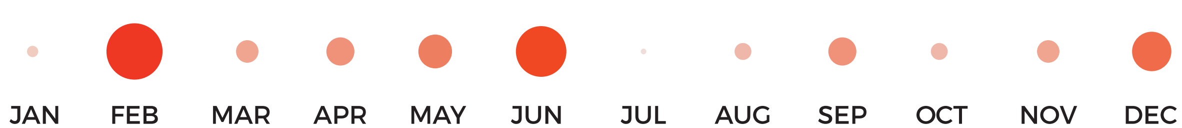
Size



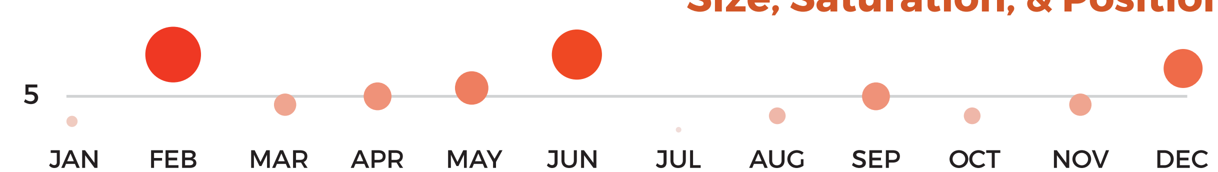
Saturation



Size & Saturation



Size, Saturation, & Position



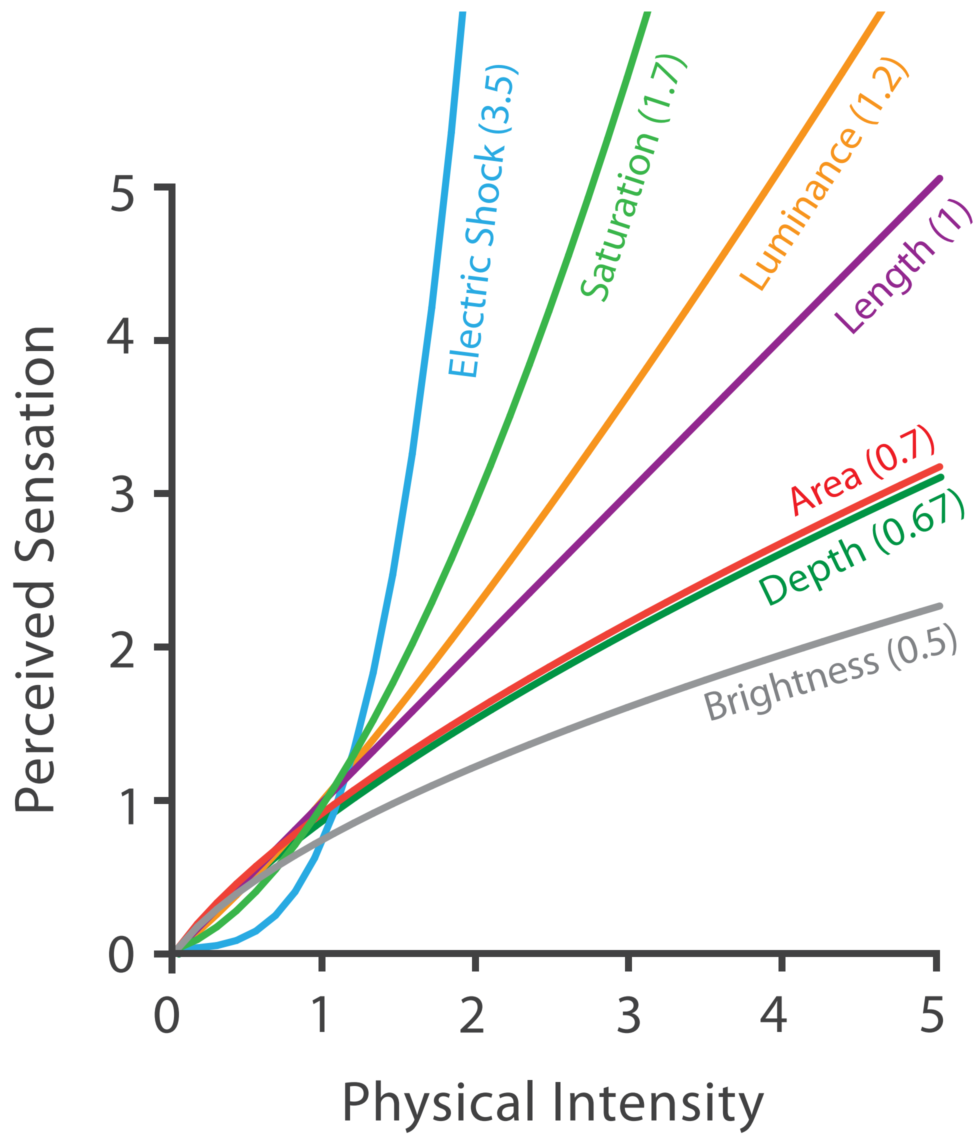
And that's just a really simple low dimensional example

Moreover, all of these visualizations encode the information, but the decode error (interpreting, comparing, …) for each graph is different

But, why?

Our perception system does not behave linearly.

Some stimuli are perceived less or more than intended.

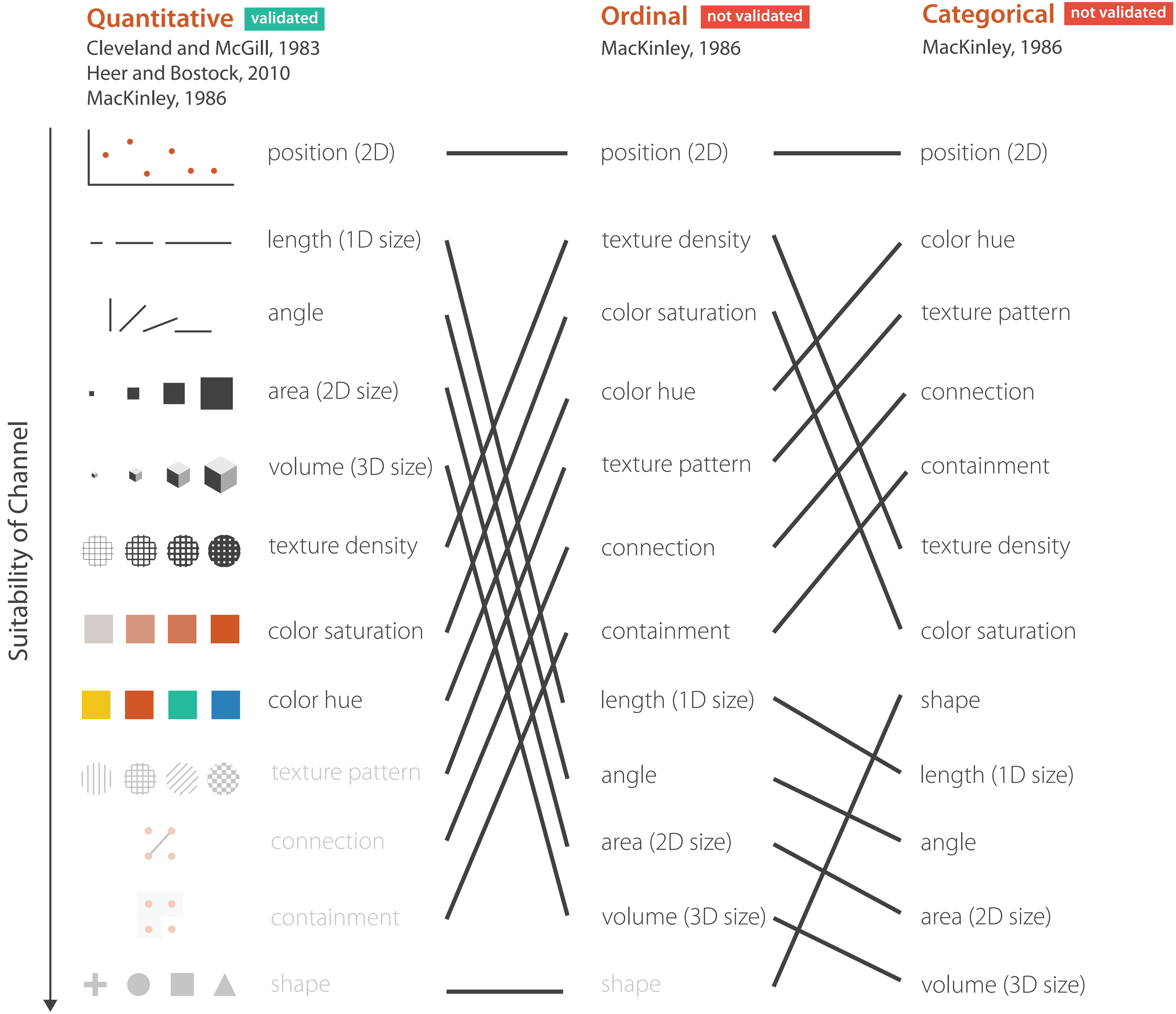


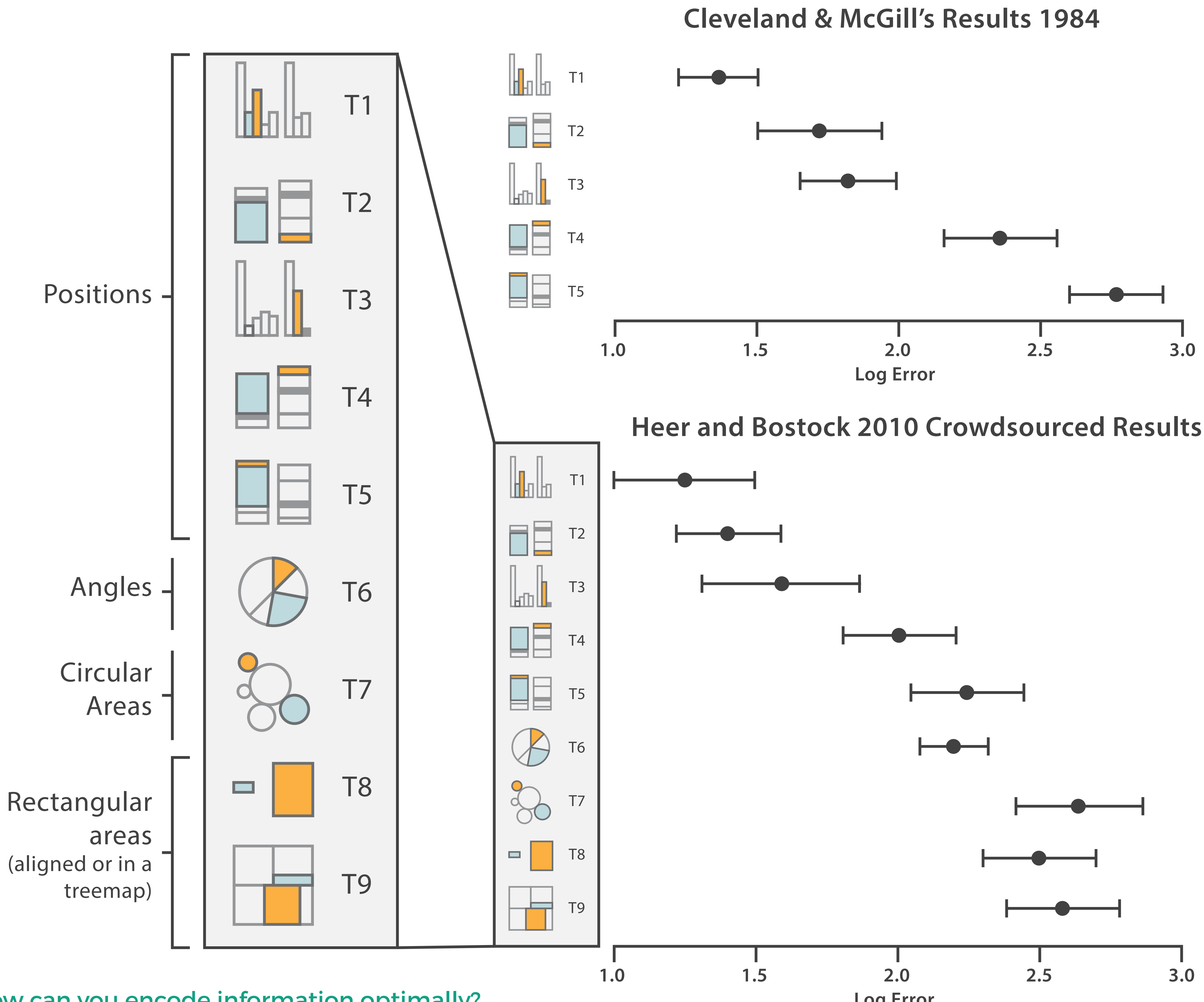
Steven's Psychophysical Power Law: $S = I^N$

Stevens, 1975

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

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

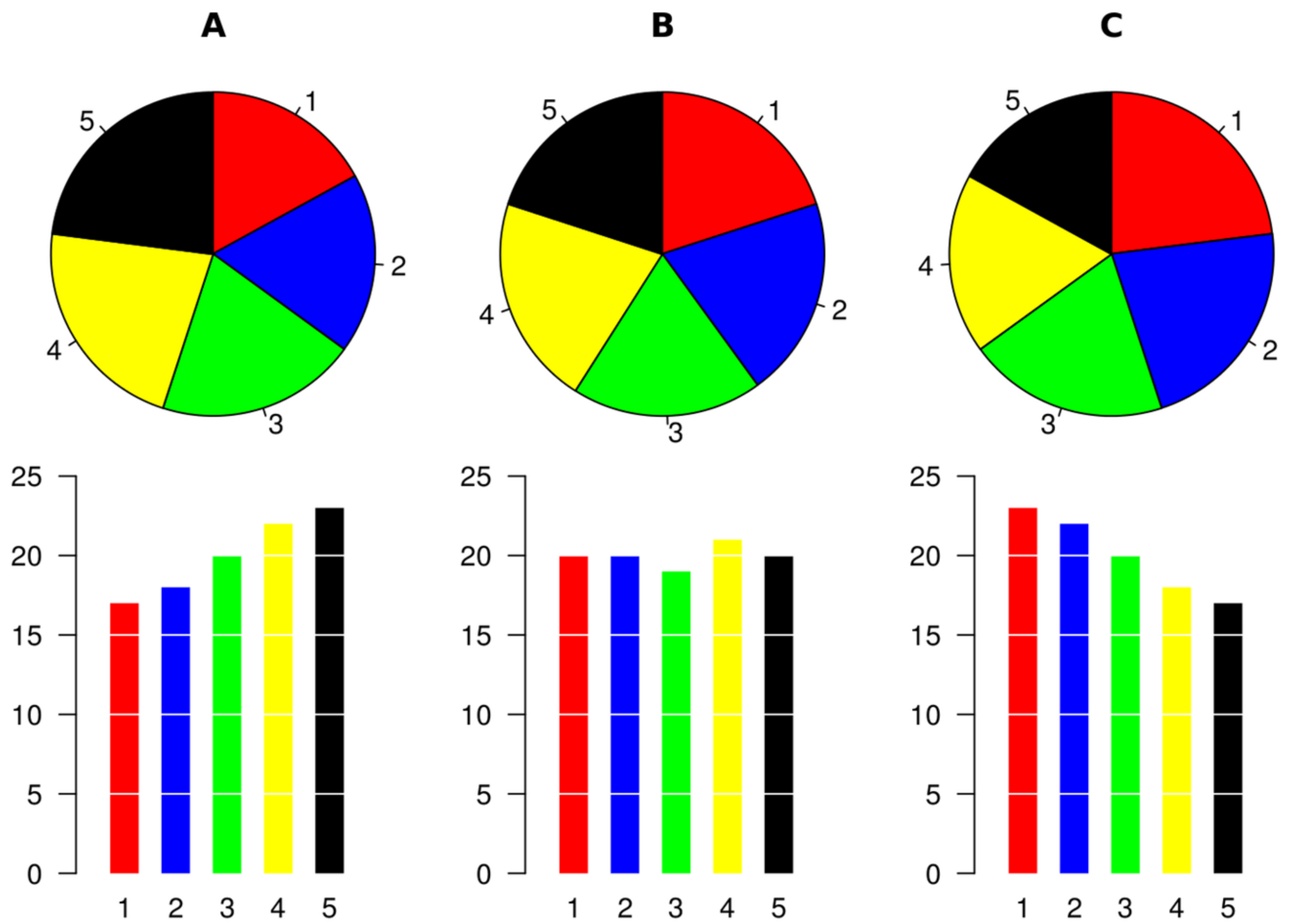

Quantitative
validated
Cleveland and McGill, 1983
Heer and Bostock, 2010
MacKinley, 1986
Ordinal
not validated
MacKinley, 1986
Categorical
not validated
MacKinley, 1986
Suitability of Channel
position (2D)
length (1D size)
angle
area (2D size)
volume (3D size)
texture density
color saturation
color hue
texture pattern
connection
containment
shape
position (2D)
texture density
color saturation
color hue
texture pattern
connection
containment
length (1D size)
angle
area (2D size)
volume (3D size)
shape
position (2D)
color hue
texture pattern
connection
containment
texture density
color saturation
shape
length (1D size)
angle
area (2D size)
volume (3D size)



How can you encode information optimally?

T6: Pie charts have also been studied in more detail recently

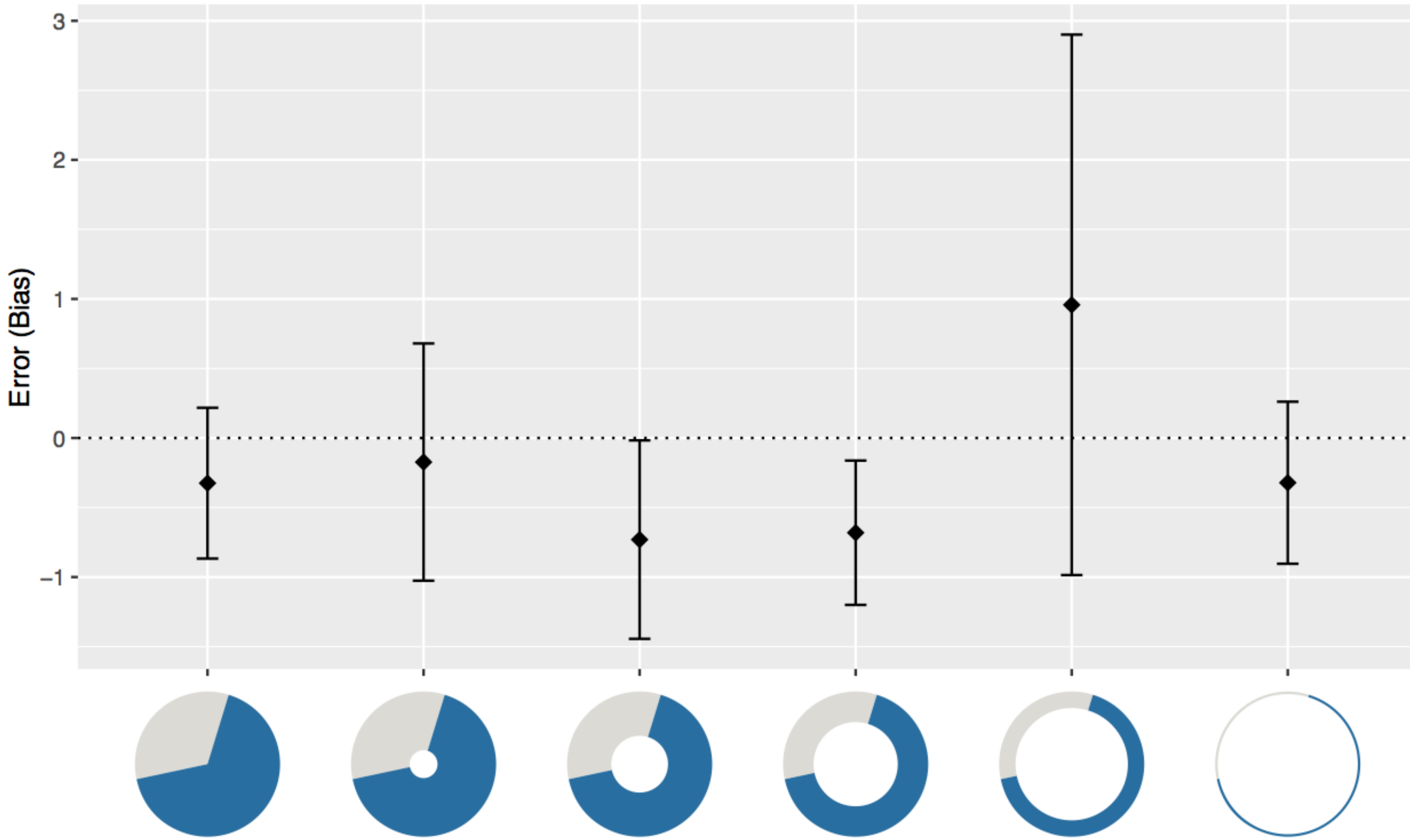
It’s quite clear that bar charts are a more effective visual encoding here than pie charts… our visual system is very good at judging lengths, but not so much at judging angles and areas.



https://commons.wikimedia.org/wiki/File:Piecharts.svg

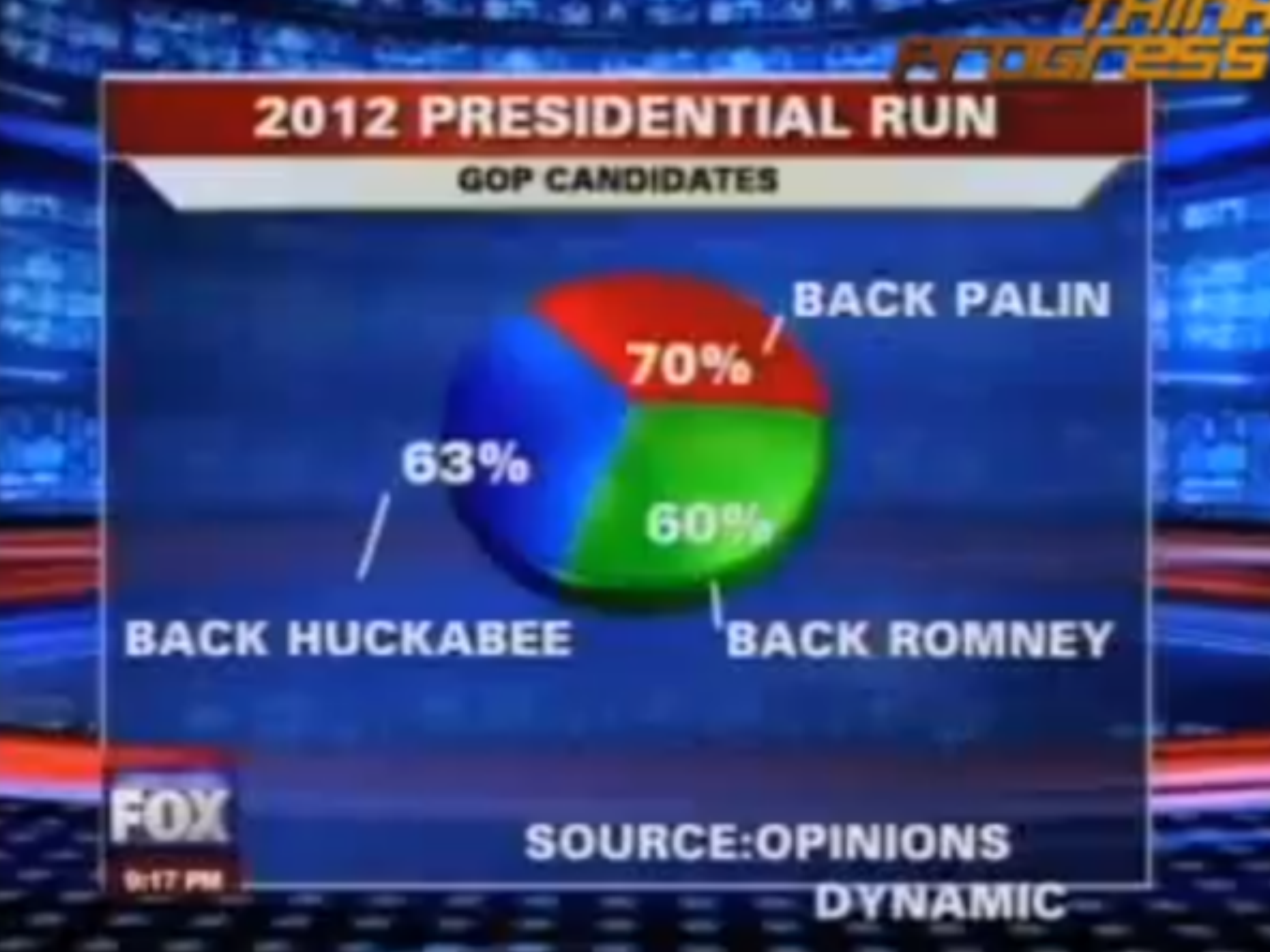
T6: Pie charts have also been studied in more detail recently

When someone reads or compares values in a pie chart, what are they doing? Comparing angles, areas, length of arc?

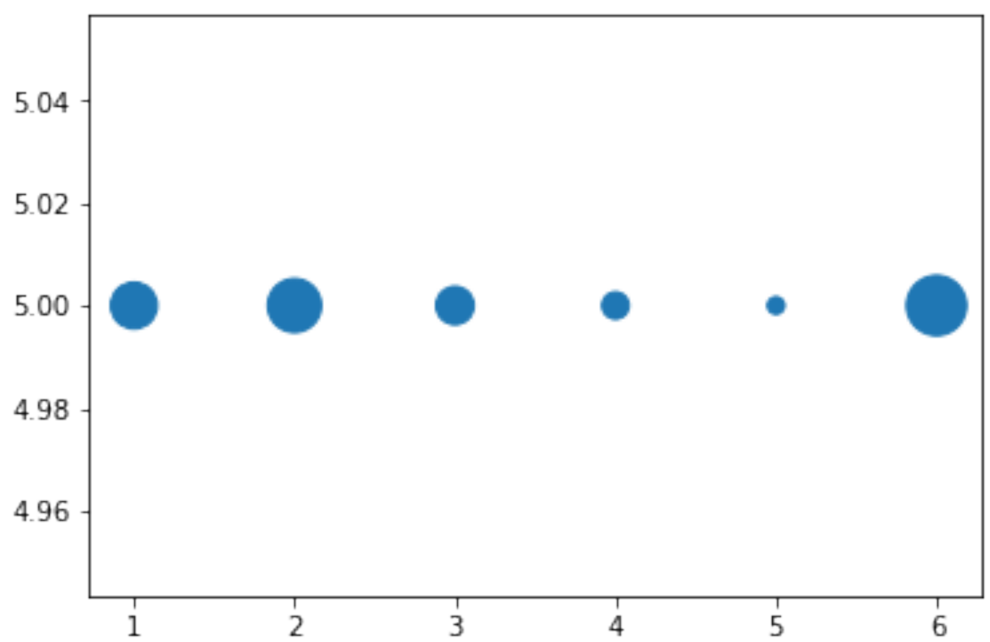


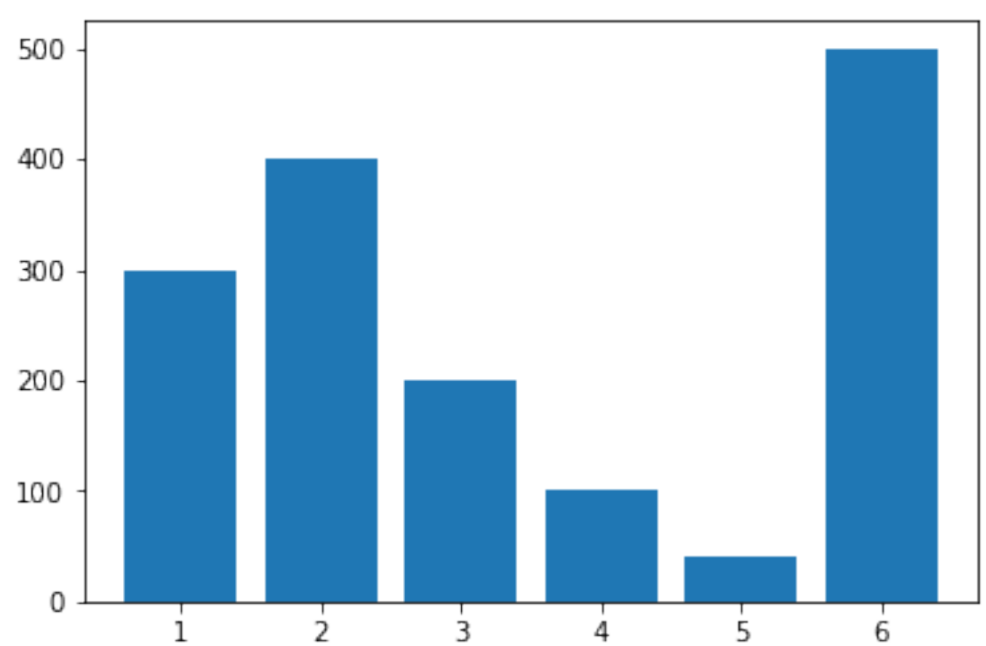
Robert Kosara and Drew Skau. 2016. **Judgment error in pie chart variations**. In Proceedings of the Eurographics: Short Papers (EuroVis '16). Eurographics Association, Goslar Germany, Germany, 91-95. DOI: https://doi.org/10.2312/eurovisshort.20161167

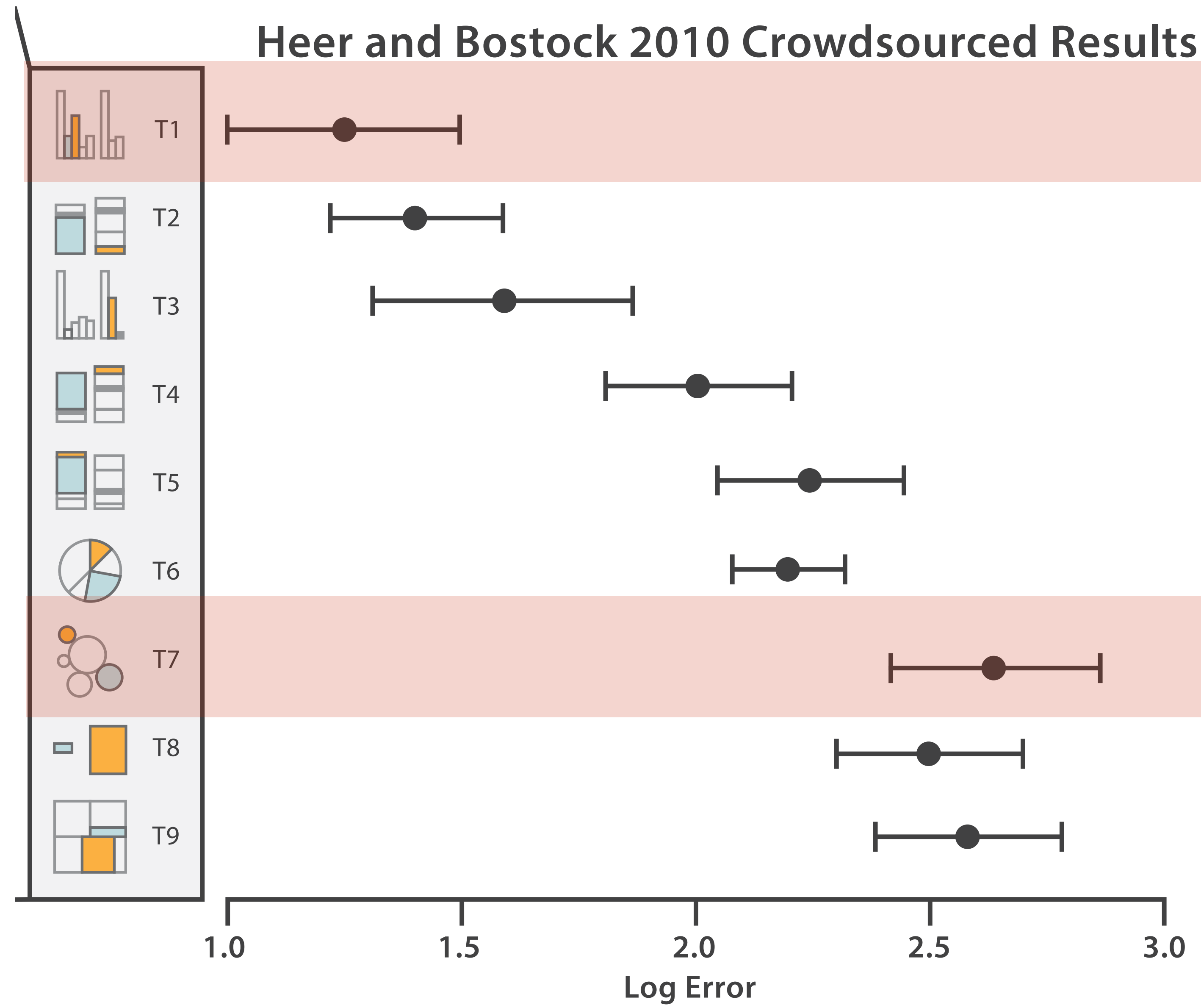
Drew Skau and Robert Kosara. 2016. **Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts**. Comput. Graph. Forum 35, 3 (June 2016), 121-130. DOI: https://doi.org/10.1111/cgf.12888


2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
SOURCE:OPINIONS
DYNAMIC

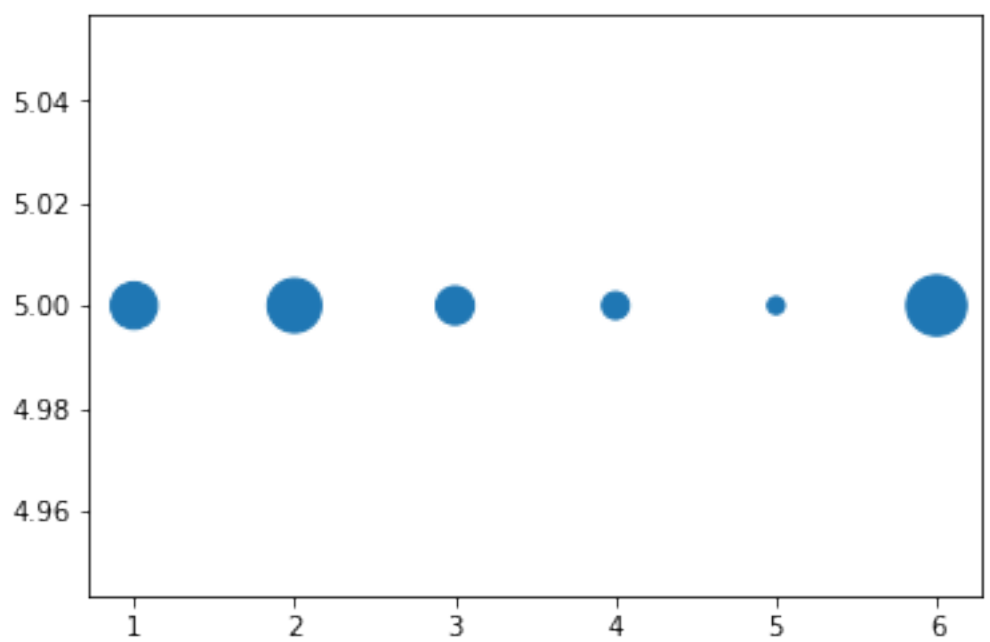
T1/T7: Bar charts are better than areas...

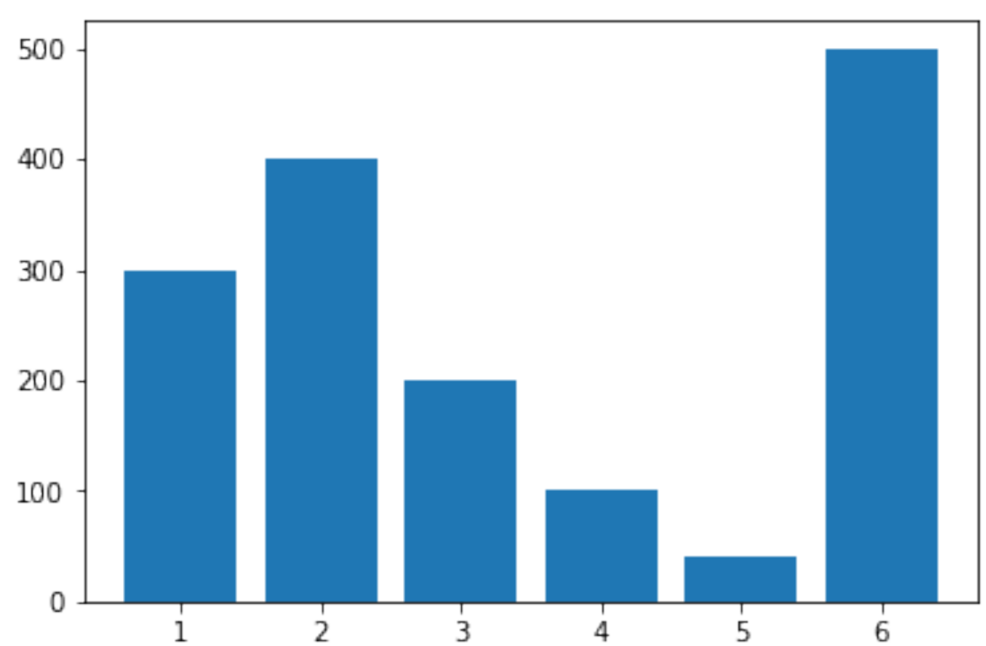




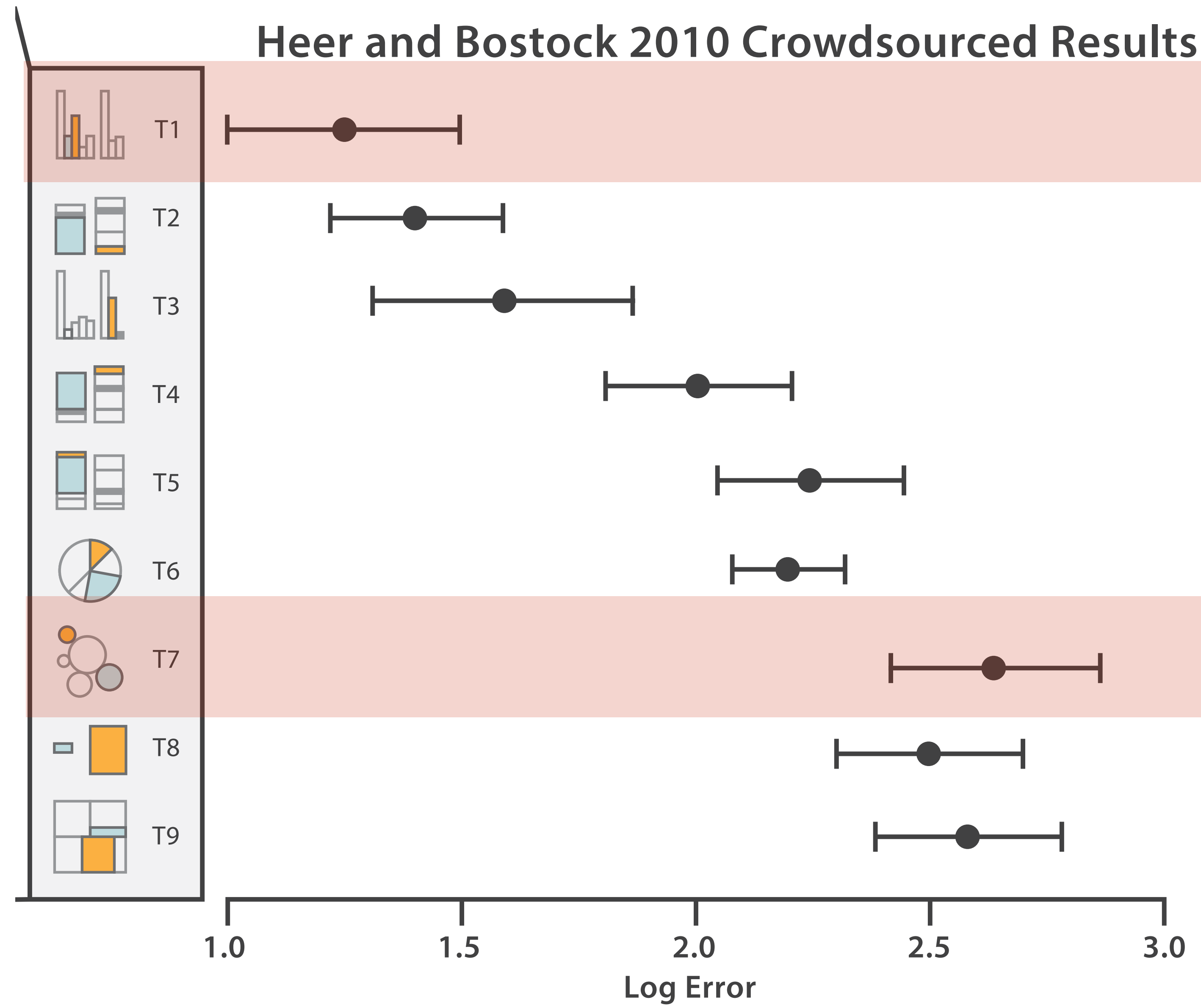


T1/T7: Bar charts are better than areas...



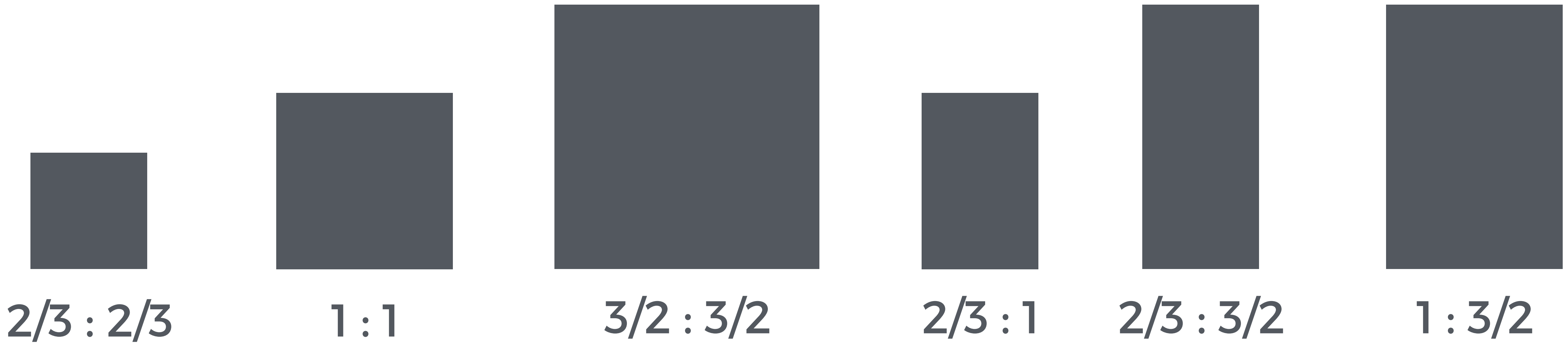


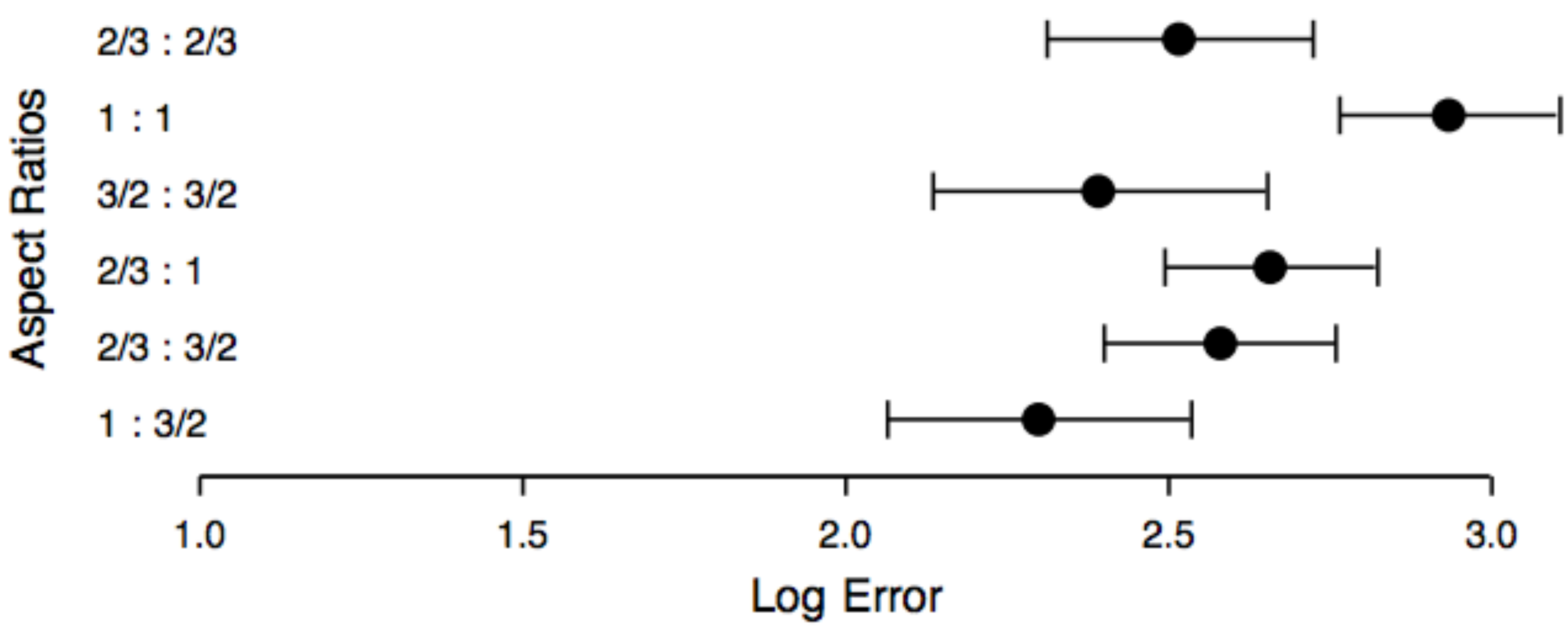
300 400 200 100 40 500



This is exactly the same data,
at the right scaling.

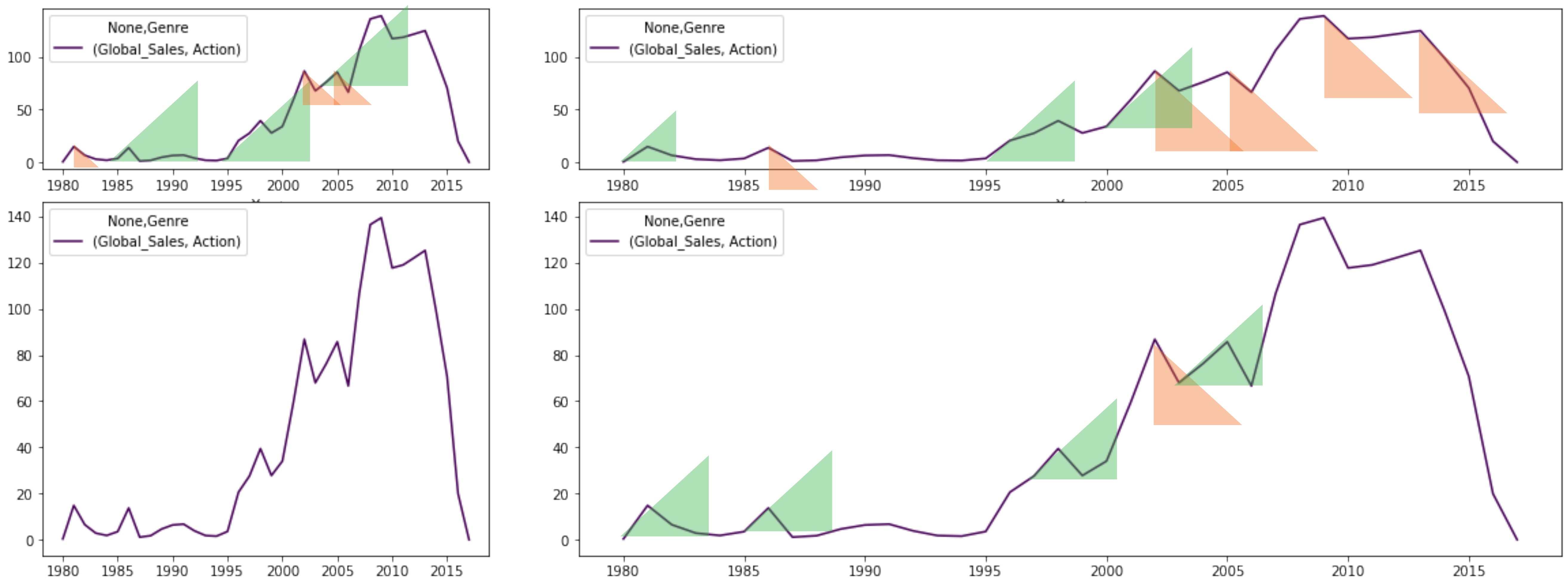
T8/T9: Different aspect ratios for rectangles also result in greater or fewer errors in estimating





Aspect ratio is important!

For line charts there is a basic guideline on optimising plot aspect ratio to have an average angle of 45 degrees from Cleveland et al, 1988.



Although, like most things, not everyone agrees with this guideline.
In this case I think it makes sense, you can decide :)

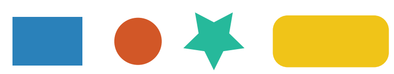
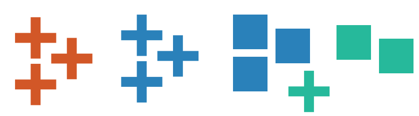
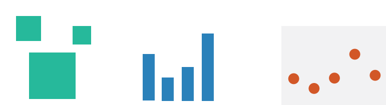
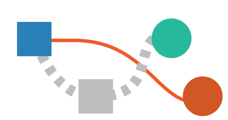
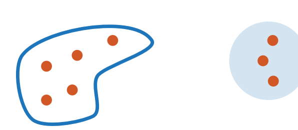
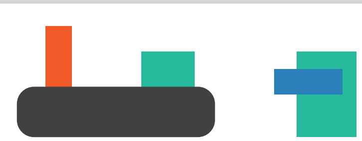
HOW

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

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

Some data has a **natural mapping** that our brains expect given certain types of data

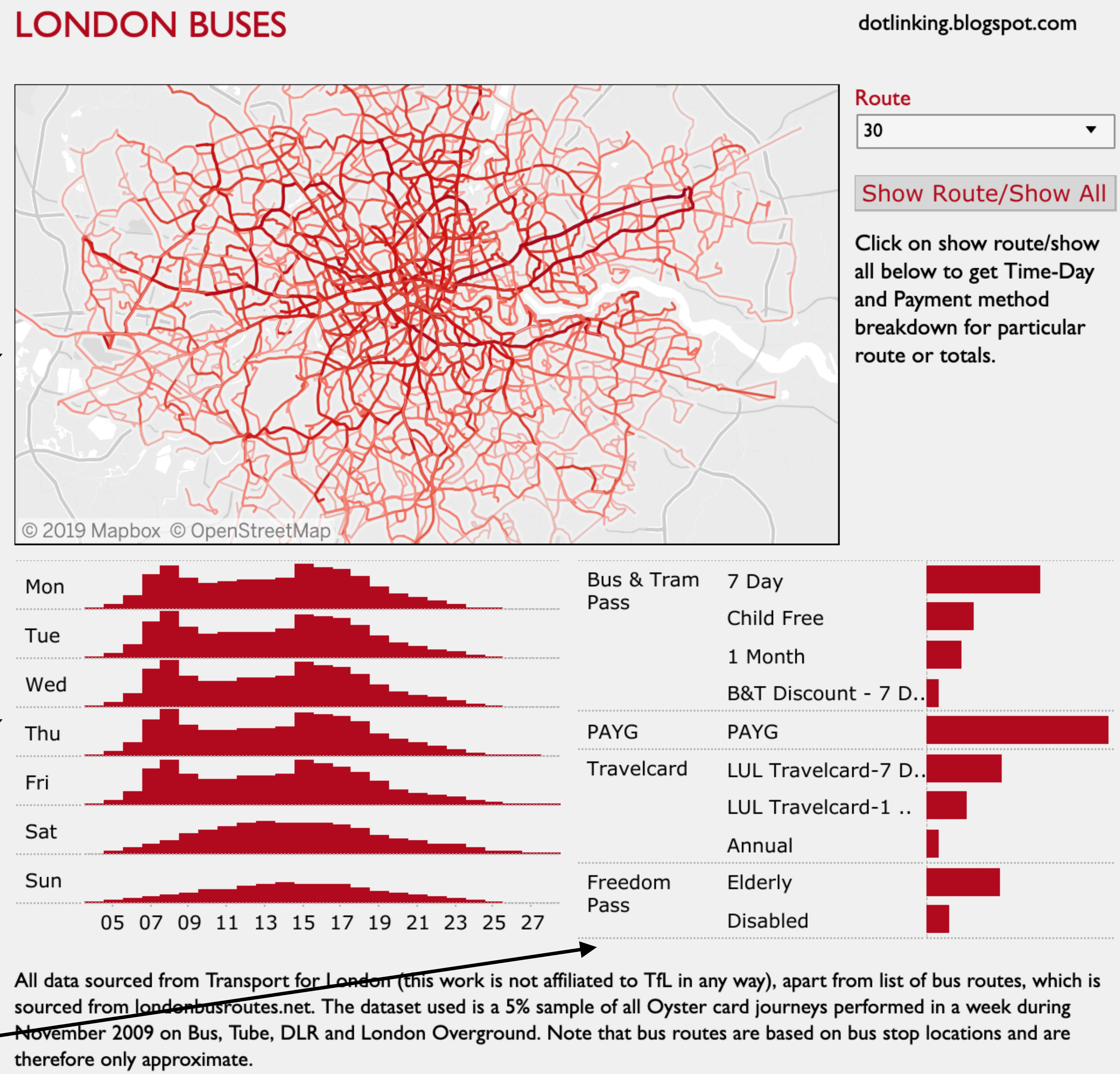
Natural Mappings

Graphical Code		Semantics
Small shapes defined by closed contour, texture, color.		Object, idea, entity, node.
Spatially ordered graphical objects.		Related information or a sequence. In a sequence the left-to-right ordering convention is borrowed from written language (English, French, etc.).
Graphical objects in proximity		Similar concepts
Graphical objects having the same shape color, or texture.		Similar concepts
Size, position or height of graphical object		Size, quantity, importance, 2D location
Shapes connected by contour		Related entities, path between entities.
Thickness of connecting contour		Strength of relationship.
Color and texture of connecting contour		Type of relationship.
Shapes enclosed by a contour, a common texture or color		Contained/related entities.
Nested/partitioned regions		Hierarchical concepts.
Attached shapes		Parts of a conceptual structure.

Locations presented on a map.

Time should be on an X-Axis

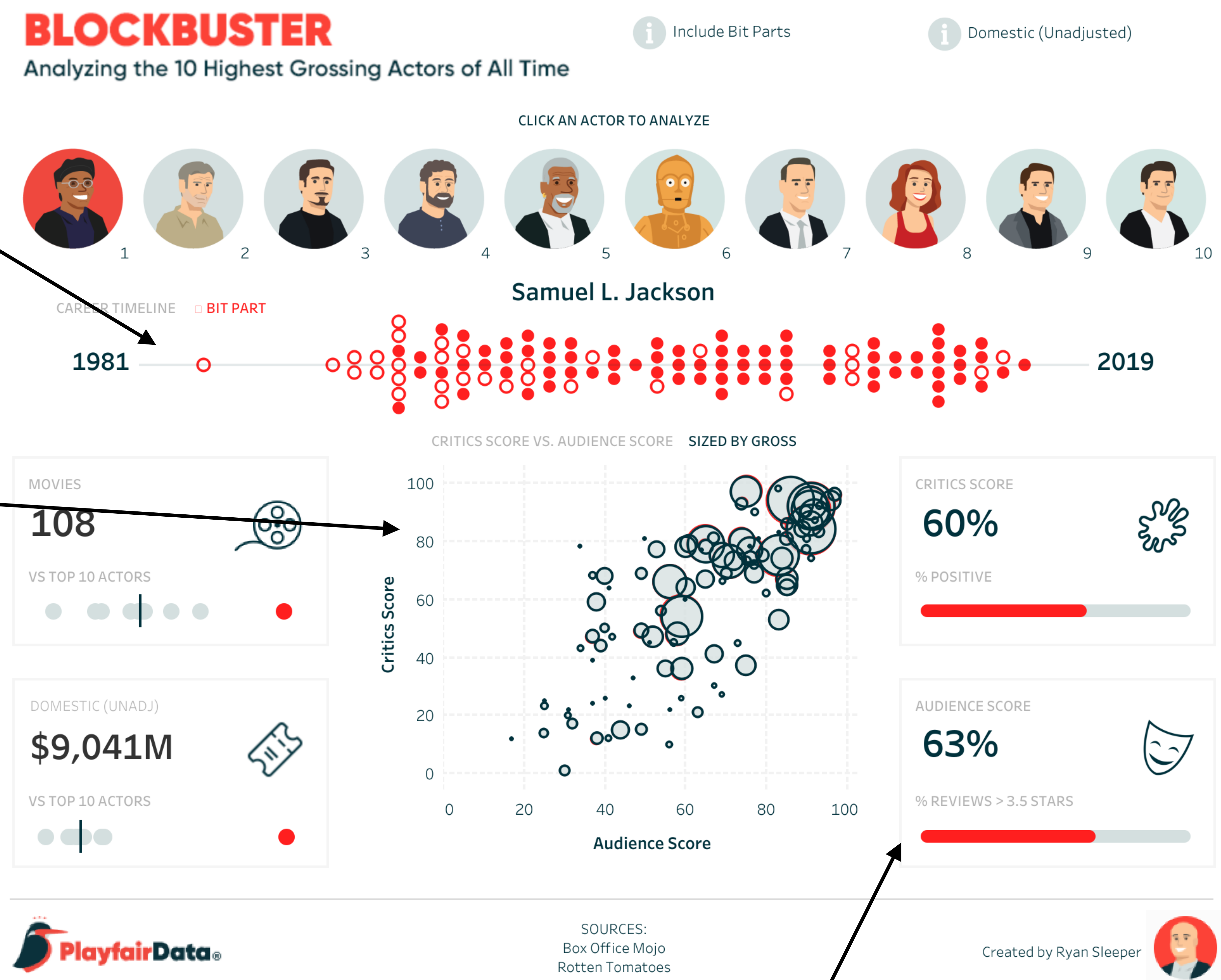
Bar charts for comparisons.



https://public.tableau.com/s/gallery/london-bus-map

Time should be on an X-Axis

Correlations as a scatter plot.



Bar charts for comparisons.

https://public.tableau.com/s/gallery/10-highest-grossing-actors-all-time?tab=featured&type=featured

HOW

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

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

Some data has a **natural mapping** that our brains expect given certain types of data

There are many intricacies of the visual system that must be considered

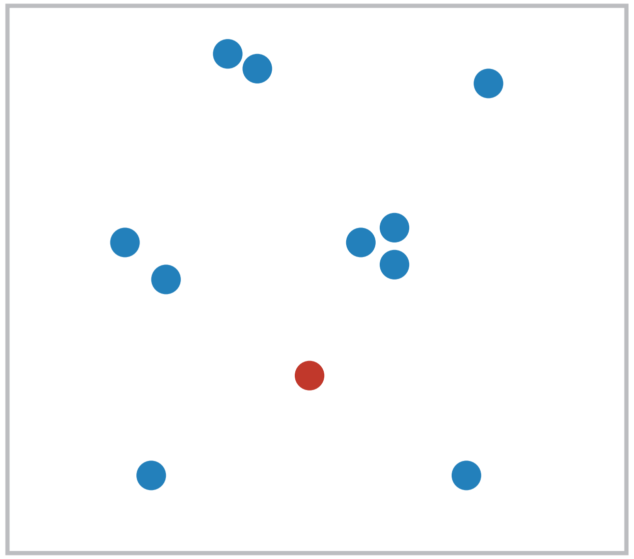
The pop-out effect

We pre-attentively process a scene, and some visual elements stand out more than others.

- Parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - speed depends on number of distractors

The pop-out effect

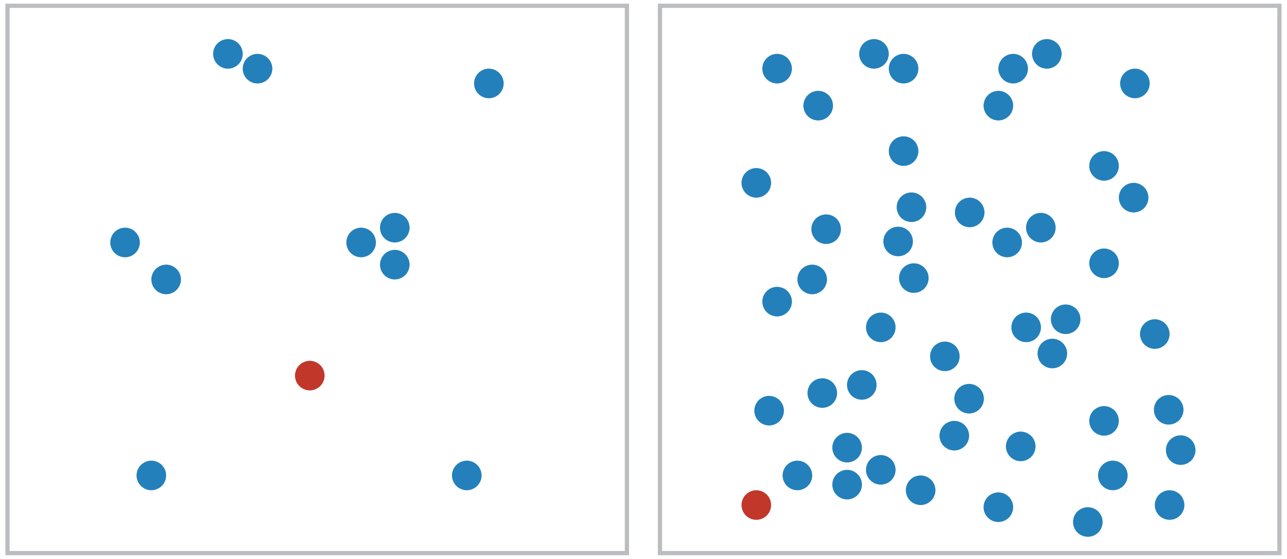
We pre-attentively process a scene, and some visual elements stand out more than others.



- Parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - speed depends on number of distractors

The pop-out effect

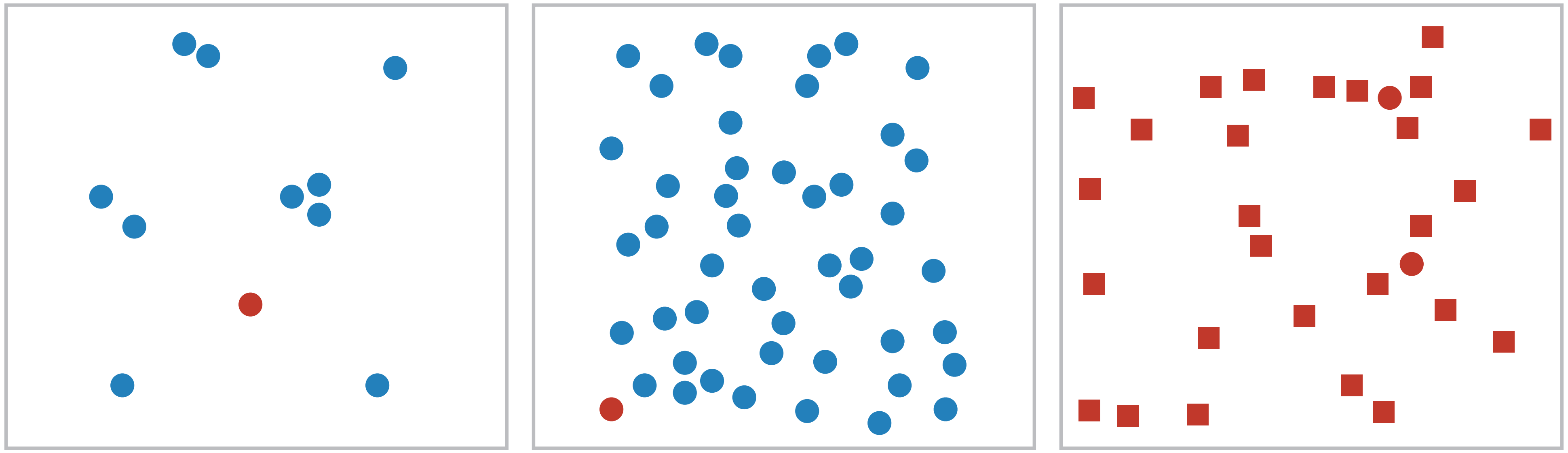
We pre-attentively process a scene, and some visual elements stand out more than others.



- Parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - speed depends on number of distractors

The pop-out effect

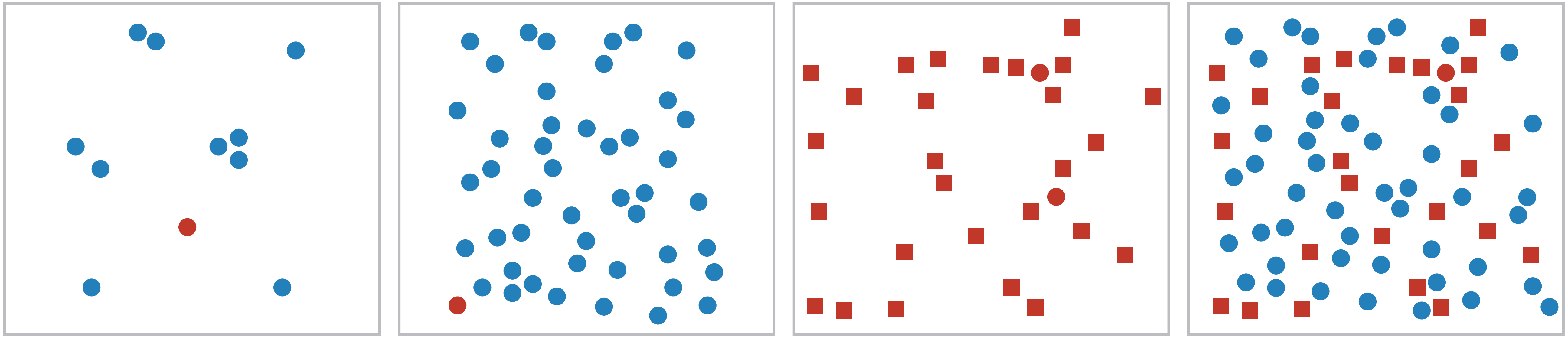
We pre-attentively process a scene, and some visual elements stand out more than others.



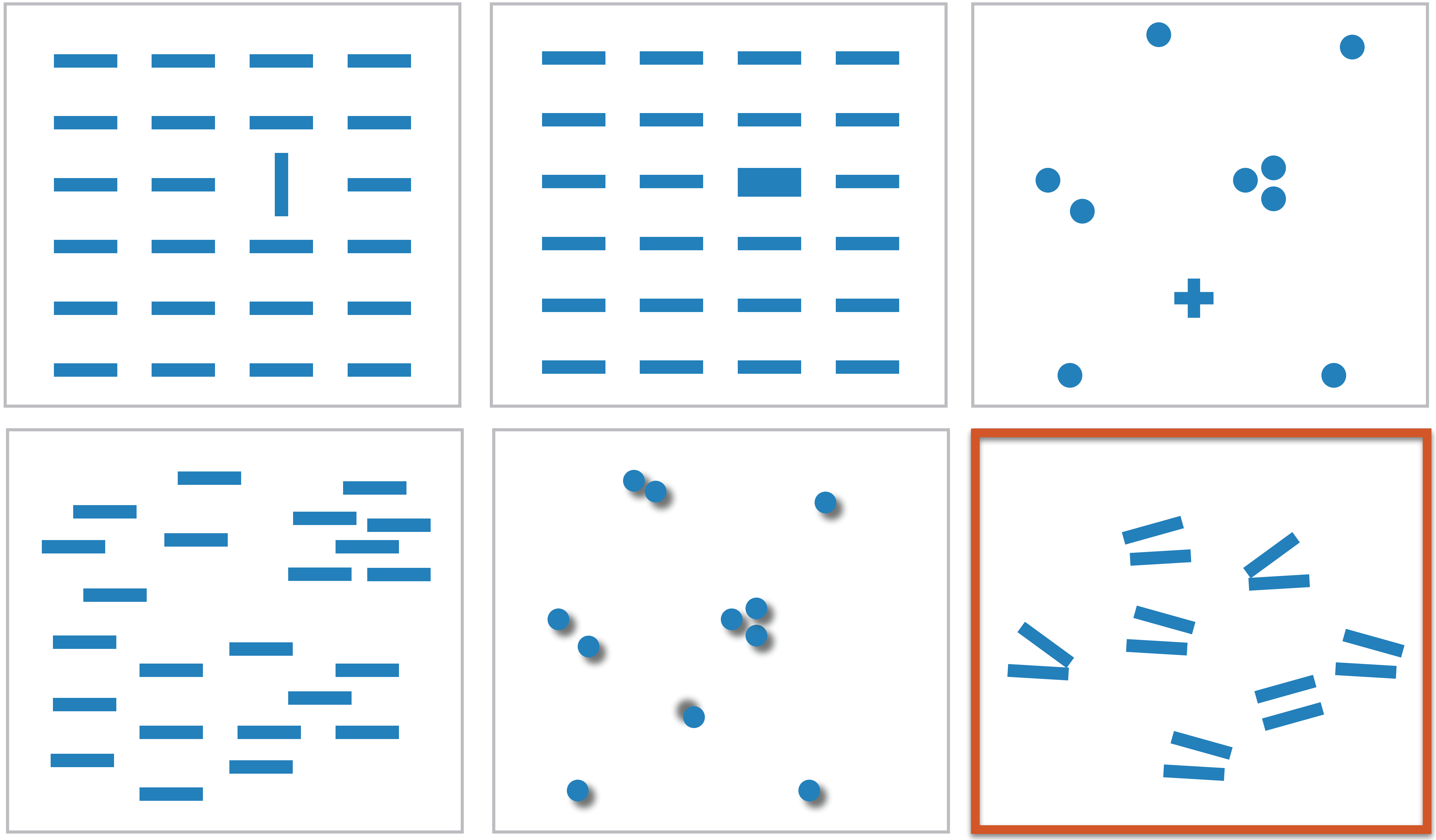
- Parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - speed depends on number of distractors

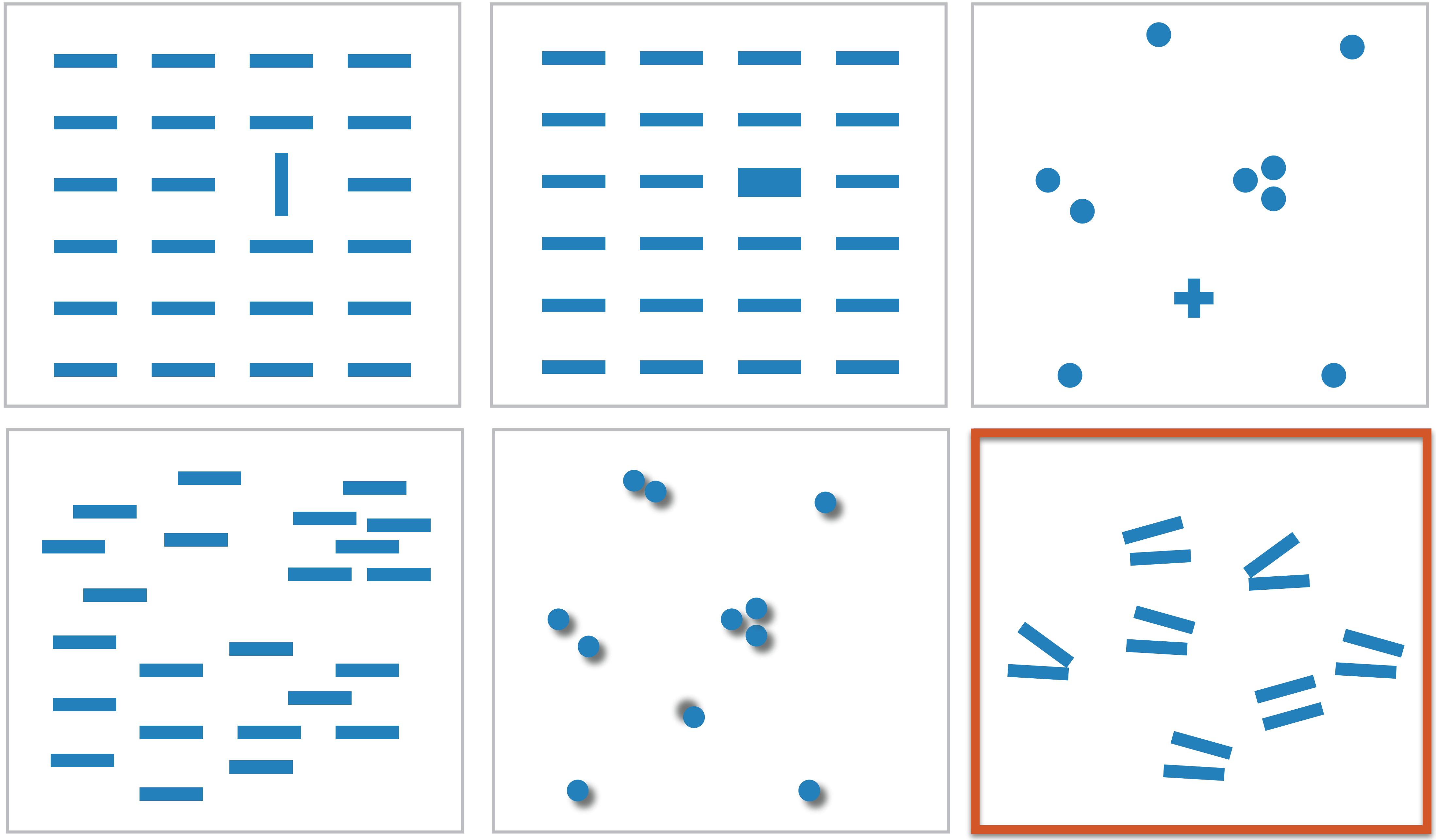
The pop-out effect

We pre-attentively process a scene, and some visual elements stand out more than others.

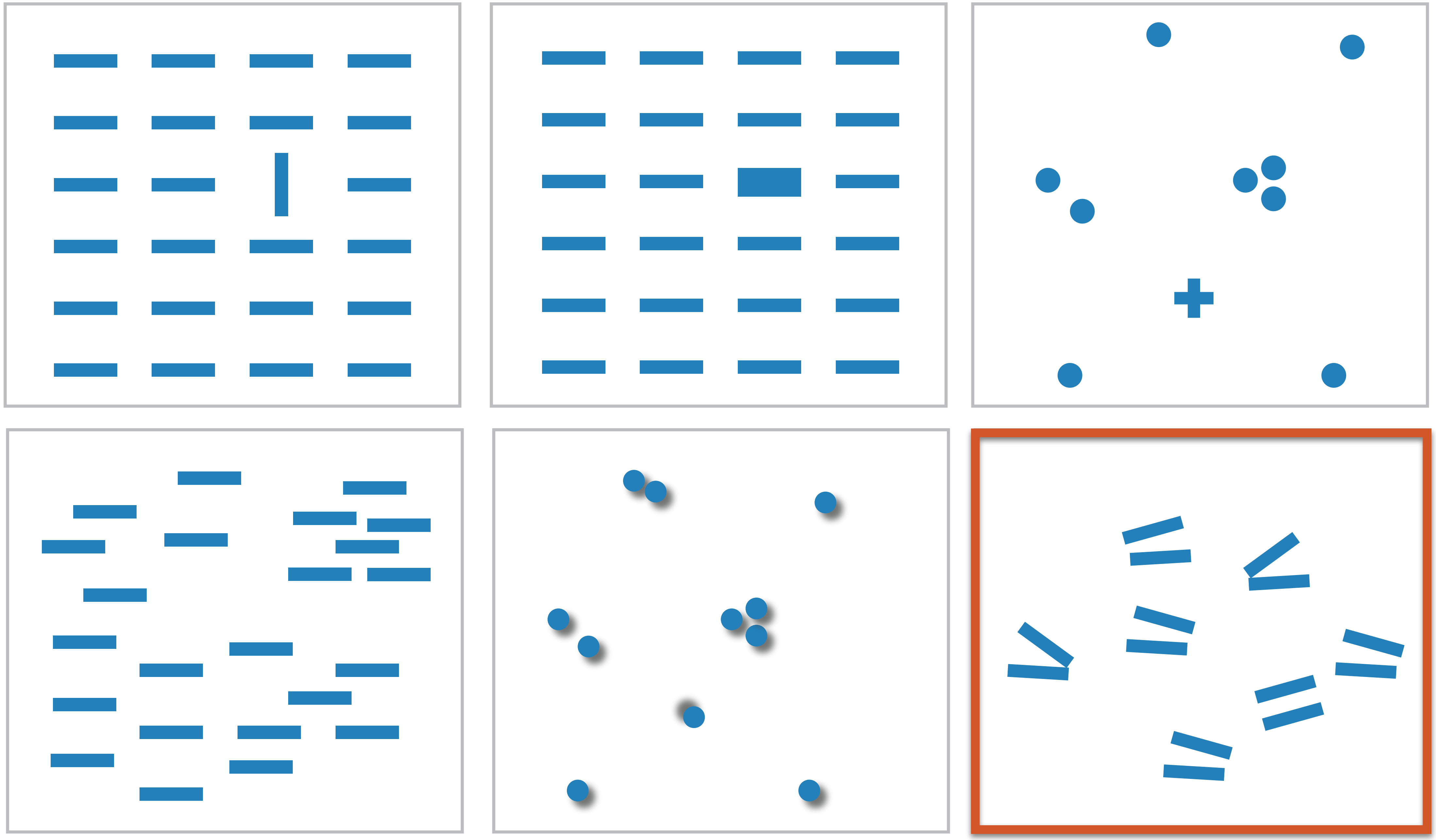


- Parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - speed depends on number of distractors



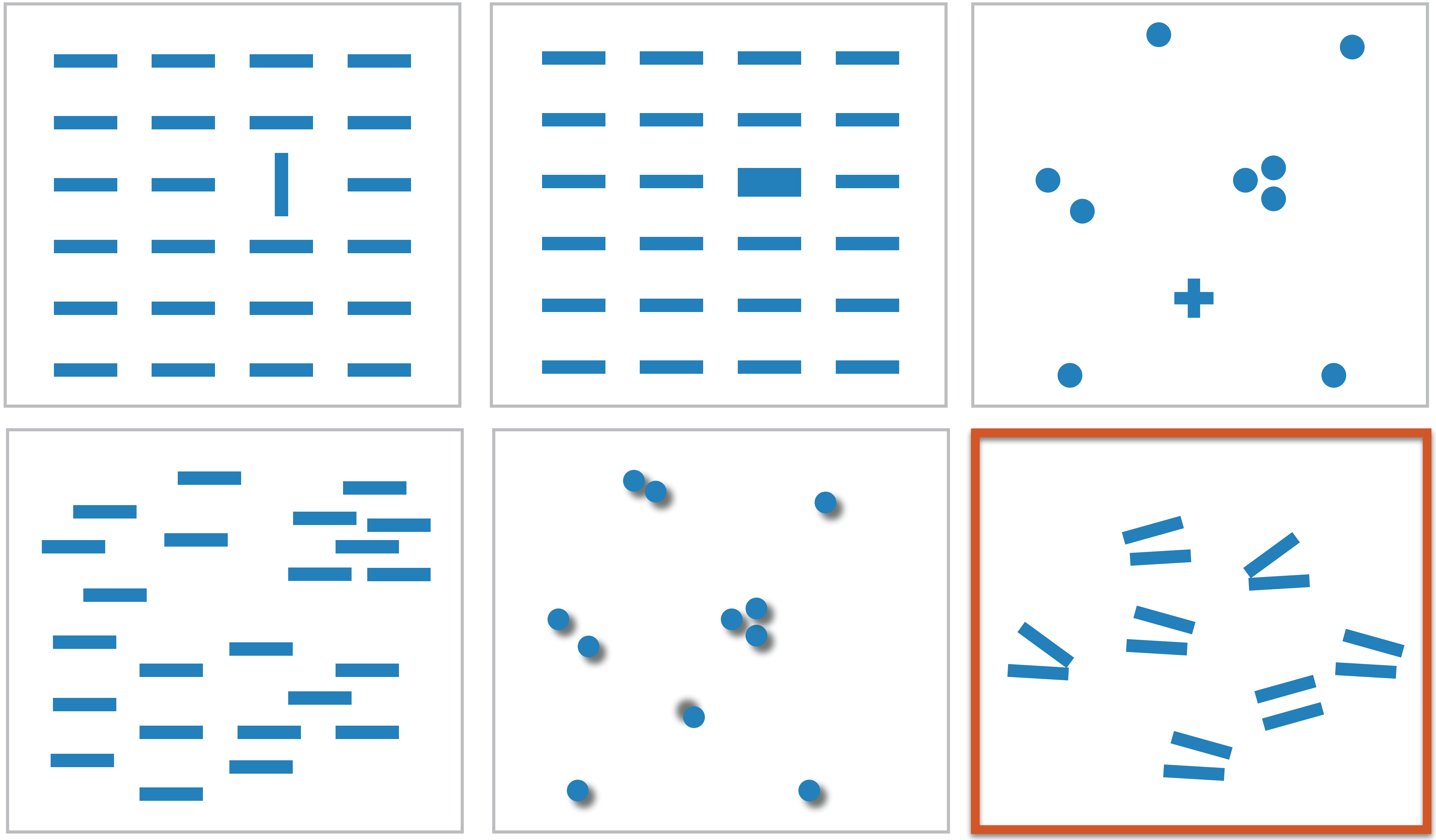


Not all exhibit the pop-out effect!



Not all exhibit the pop-out effect!

Parallel line pairs do not pop out from tilted pairs...



Not all exhibit the pop-out effect!

Parallel line pairs do not pop out from tilted pairs...

And not all visual channels pop out as quickly as other. E.g. colour is always on top.

The pop-out effect

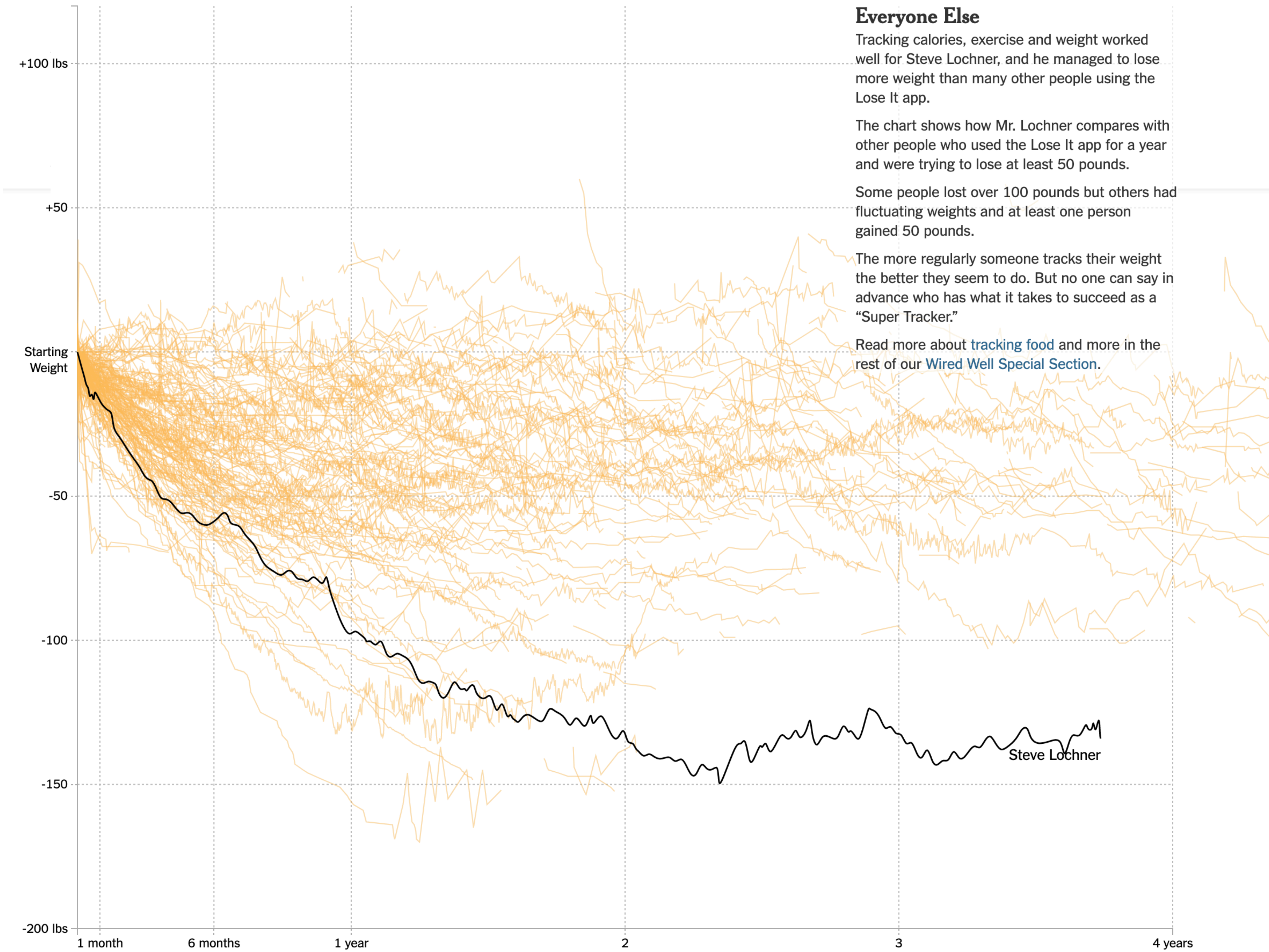
3	3	0	3	0	1	8	7	6	8	2	1	4	0	3	8	3	7	7	2	0	5	2	3	2	7	0	2	0
7	1	4	6	0	2	1	3	2	7	6	0	2	5	6	3	2	5	7	6	3	3	0	2	0	3	0	7	2
8	7	5	7	2	8	3	8	7	7	8	2	0	7	7	5	2	3	1	1	5	6	3	8	4	7	8	2	0
0	5	0	5	1	6	1	7	5	6	8	0	4	4	6	7	4	7	1	4	0	0	8	4	4	3	0	3	2
2	4	3	1	3	5	4	9	5	0	7	6	0	7	4	3	1	8	2	7	3	4	6	0	2	4	8	2	3
8	6	2	2	6	5	4	6	7	0	7	6	0	0	3	9	0	2	4	7	1	7	2	3	3	5	8	7	0
0	8	4	5	1	3	1	7	6	4	5	4	1	2	4	5	3	3	5	4	9	6	7	7	6	3	4	2	5
4	7	7	0	2	2	0	1	1	7	7	7	0	2	6	6	4	7	5	8	6	1	4	3	7	8	5	4	6
4	3	6	6	4	6	6	2	8	4	8	5	3	7	8	8	1	3	8	5	4	5	7	4	0	3	2	8	4
5	5	0	3	5	3	5	3	8	3	2	3	8	2	3	1	6	2	7	2	4	6	3	6	4	4	3	2	5
4	4	0	2	1	7	2	4	4	7	4	1	9	2	4	5	2	5	0	4	0	0	5	3	6	3	3	6	7
7	4	6	6	8	7	5	7	9	2	0	2	8	8	8	8	3	2	4	2	6	4	0	4	6	3	7	2	1
0	1	7	1	5	9	1	4	2	8	7	3	7	1	4	5	1	8	7	8	0	5	1	7	0	5	8	8	1
2	8	5	2	1	2	8	7	7	6	2	5	6	2	6	4	1	5	1	6	1	2	1	1	0	5	6	4	0
2	1	1	7	7	2	0	0	1	8	7	0	2	9	0	2	8	5	7	8	4	6	0	6	5	0	7	1	2
0	5	2	4	1	5	3	3	1	5	5	1	4	0	1	6	4	3	3	9	8	8	3	4	6	8	4	8	6
7	3	7	5	2	4	0	2	7	6	3	8	5	5	4	5	8	8	7	5	5	6	5	6	7	9	7	7	4
0	3	2	8	1	4	4	6	0	8	2	3	0	1	3	4	6	2	0	5	7	7	3	6	1	8	7	3	5
4	4	8	3	3	3	5	0	1	0	3	8	6	3	2	0	5	0	6	1	3	3	4	3	6	1	5	8	6
1	0	2	2	7	6	3	3	0	8	8	0	3	1	8	8	1	2	1	7	5	2	9	3	5	8	3	2	5

The pop-out effect

3	3	0	3	0	1	8	7	6	8	2	1	4	0	3	8	3	7	7	2	0	5	2	3	2	7	0	2	0
7	1	4	6	0	2	1	3	2	7	6	0	2	5	6	3	2	5	7	6	3	3	0	2	0	3	0	7	2
8	7	5	7	2	8	3	8	7	7	8	2	0	7	7	5	2	3	1	1	5	6	3	8	4	7	8	2	0
0	5	0	5	1	6	1	7	5	6	8	0	4	4	6	7	4	7	1	4	0	0	8	4	4	3	0	3	2
2	4	3	1	3	5	4	9	5	0	7	6	0	7	4	3	1	8	2	7	3	4	6	0	2	4	8	2	3
8	6	2	2	6	5	4	6	7	0	7	6	0	0	3	9	0	2	4	7	1	7	2	3	3	5	8	7	0
0	8	4	5	1	3	1	7	6	4	5	4	1	2	4	5	3	3	5	4	9	6	7	7	6	3	4	2	5
4	7	7	0	2	2	0	1	1	7	7	7	0	2	6	6	4	7	5	8	6	1	4	3	7	8	5	4	6
4	3	6	6	4	6	6	2	8	4	8	5	3	7	8	8	1	3	8	5	4	5	7	4	0	3	2	8	4
5	5	0	3	5	3	5	3	8	3	2	3	8	2	3	1	6	2	7	2	4	6	3	6	4	4	3	2	5
4	4	0	2	1	7	2	4	4	7	4	1	9	2	4	5	2	5	0	4	0	0	5	3	6	3	3	6	7
7	4	6	6	8	7	5	7	9	2	0	2	8	8	8	8	3	2	4	2	6	4	0	4	6	3	7	2	1
0	1	7	1	5	9	1	4	2	8	7	3	7	1	4	5	1	8	7	8	0	5	1	7	0	5	8	8	1
2	8	5	2	1	2	8	7	7	6	2	5	6	2	6	4	1	5	1	6	1	2	1	1	0	5	6	4	0
2	1	1	7	7	2	0	0	1	8	7	0	2	9	0	2	8	5	7	8	4	6	0	6	5	0	7	1	2
0	5	2	4	1	5	3	3	1	5	5	1	4	0	1	6	4	3	3	9	8	8	3	4	6	8	4	8	6
7	3	7	5	2	4	0	2	7	6	3	8	5	5	4	5	8	8	7	5	5	6	5	6	7	9	7	7	4
0	3	2	8	1	4	4	6	0	8	2	3	0	1	3	4	6	2	0	5	7	7	3	6	1	8	7	3	5
4	4	8	3	3	3	5	0	1	0	3	8	6	3	2	0	5	0	6	1	3	3	4	3	6	1	5	8	6
1	0	2	2	7	6	3	3	0	8	8	0	3	1	8	8	1	2	1	7	5	2	9	3	5	8	3	2	5

The pop-out effect

3	3	0	3	0	1	8	7	6	8	2	1	4	0	3	8	3	7	7	2	0	5	2	3	2	7	0	2	0
7	1	4	6	0	2	1	3	2	7	6	0	2	5	6	3	2	5	7	6	3	3	0	2	0	3	0	7	2
8	7	5	7	2	8	3	8	7	7	8	2	0	7	7	5	2	3	1	1	5	6	3	8	4	7	8	2	0
0	5	0	5	1	6	1	7	5	6	8	0	4	4	6	7	4	7	1	4	0	0	8	4	4	3	0	3	2
2	4	3	1	3	5	4	9	5	0	7	6	0	7	4	3	1	8	2	7	3	4	6	0	2	4	8	2	3
8	6	2	2	6	5	4	6	7	0	7	6	0	0	3	9	0	2	4	7	1	7	2	3	3	5	8	7	0
0	8	4	5	1	3	1	7	6	4	5	4	1	2	4	5	3	3	5	4	9	6	7	7	6	3	4	2	5
4	7	7	0	2	2	0	1	1	7	7	7	0	2	6	6	4	7	5	8	6	1	4	3	7	8	5	4	6
4	3	6	6	4	6	6	2	8	4	8	5	3	7	8	8	1	3	8	5	4	5	7	4	0	3	2	8	4
5	5	0	3	5	3	5	3	8	3	2	3	8	2	3	1	6	2	7	2	4	6	3	6	4	4	3	2	5
4	4	0	2	1	7	2	4	4	7	4	1	9	2	4	5	2	5	0	4	0	0	5	3	6	3	3	6	7
7	4	6	6	8	7	5	7	9	2	0	2	8	8	8	8	3	2	4	2	6	4	0	4	6	3	7	2	1
0	1	7	1	5	9	1	4	2	8	7	3	7	1	4	5	1	8	7	8	0	5	1	7	0	5	8	8	1
2	8	5	2	1	2	8	7	7	6	2	5	6	2	6	4	1	5	1	6	1	2	1	1	0	5	6	4	0
2	1	1	7	7	2	0	0	1	8	7	0	2	9	0	2	8	5	7	8	4	6	0	6	5	0	7	1	2
0	5	2	4	1	5	3	3	1	5	5	1	4	0	1	6	4	3	3	9	8	8	3	4	6	8	4	8	6
7	3	7	5	2	4	0	2	7	6	3	8	5	5	4	5	8	8	7	5	5	6	5	6	7	9	7	7	4
0	3	2	8	1	4	4	6	0	8	2	3	0	1	3	4	6	2	0	5	7	7	3	6	1	8	7	3	5
4	4	8	3	3	3	5	0	1	0	3	8	6	3	2	0	5	0	6	1	3	3	4	3	6	1	5	8	6
1	0	2	2	7	6	3	3	0	8	8	0	3	1	8	8	1	2	1	7	5	2	9	3	5	8	3	2	5

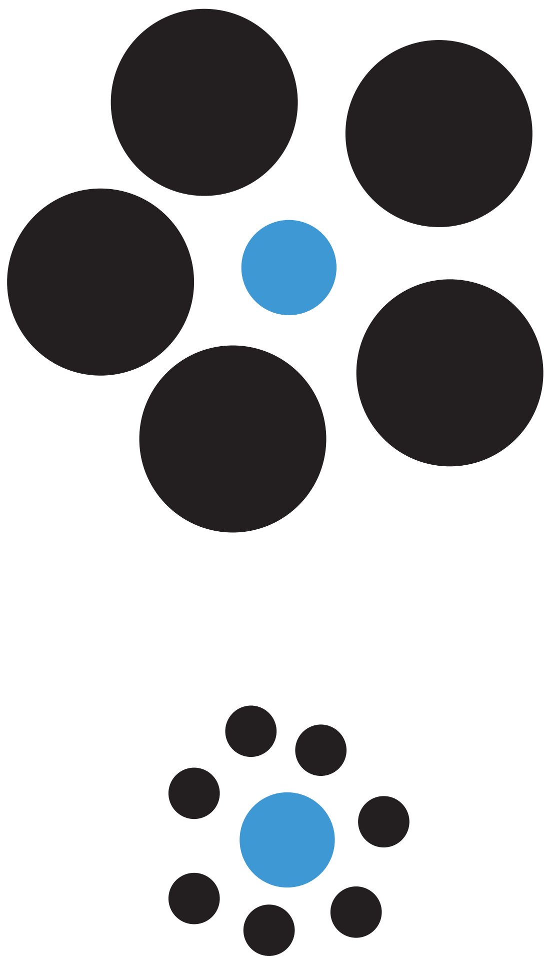


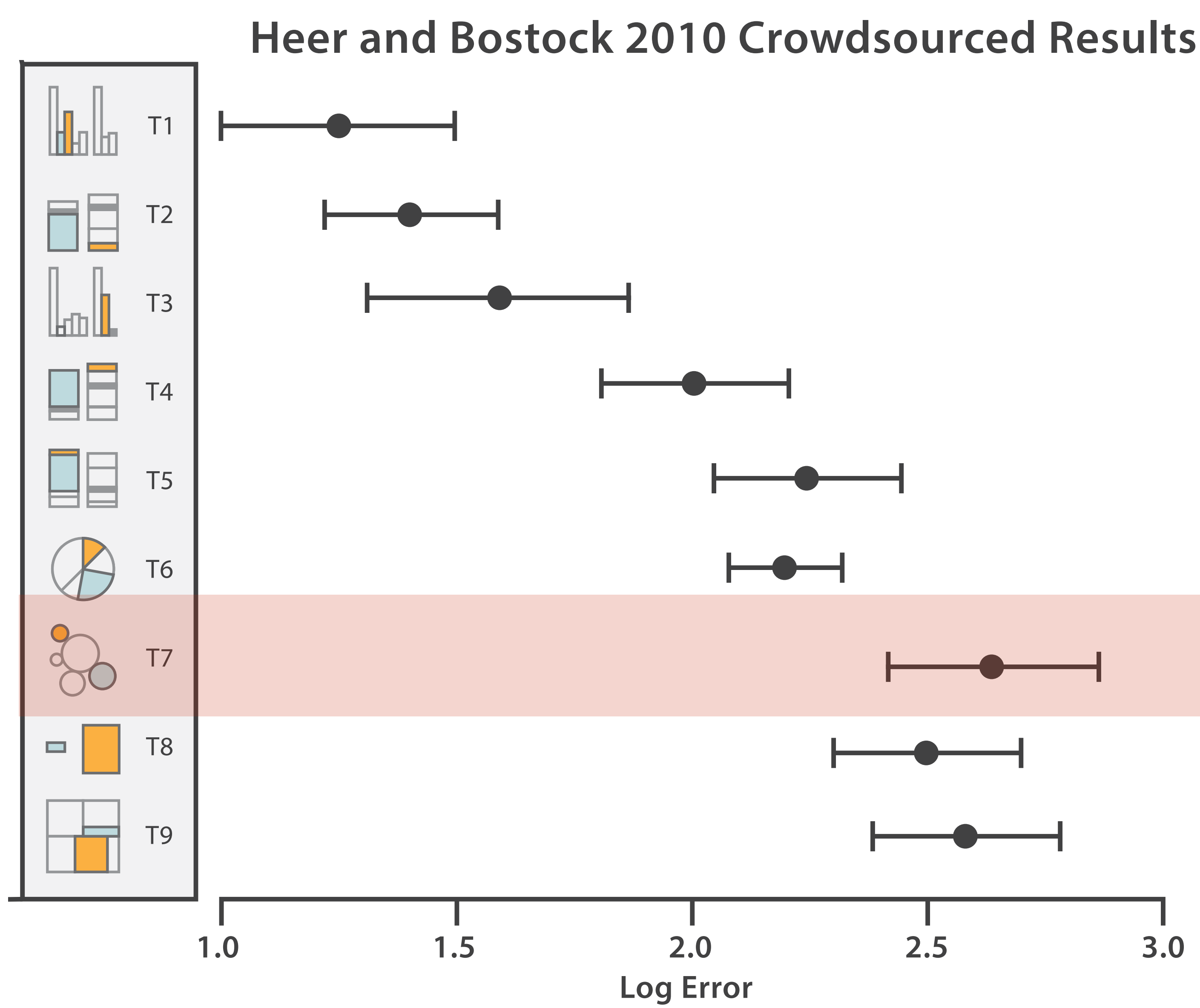
Check out https://www.nytimes.com/interactive/2015/11/17/health/wiredwell-food-diary-super-tracker.html - beautiful storytelling using visualization and annotations.

Relative Comparison

Relative Comparison

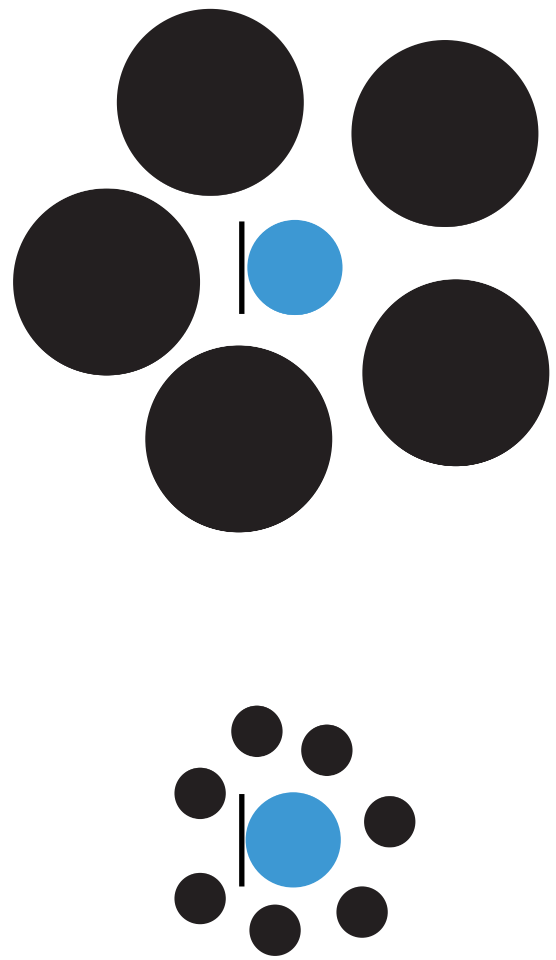
Relative Comparison

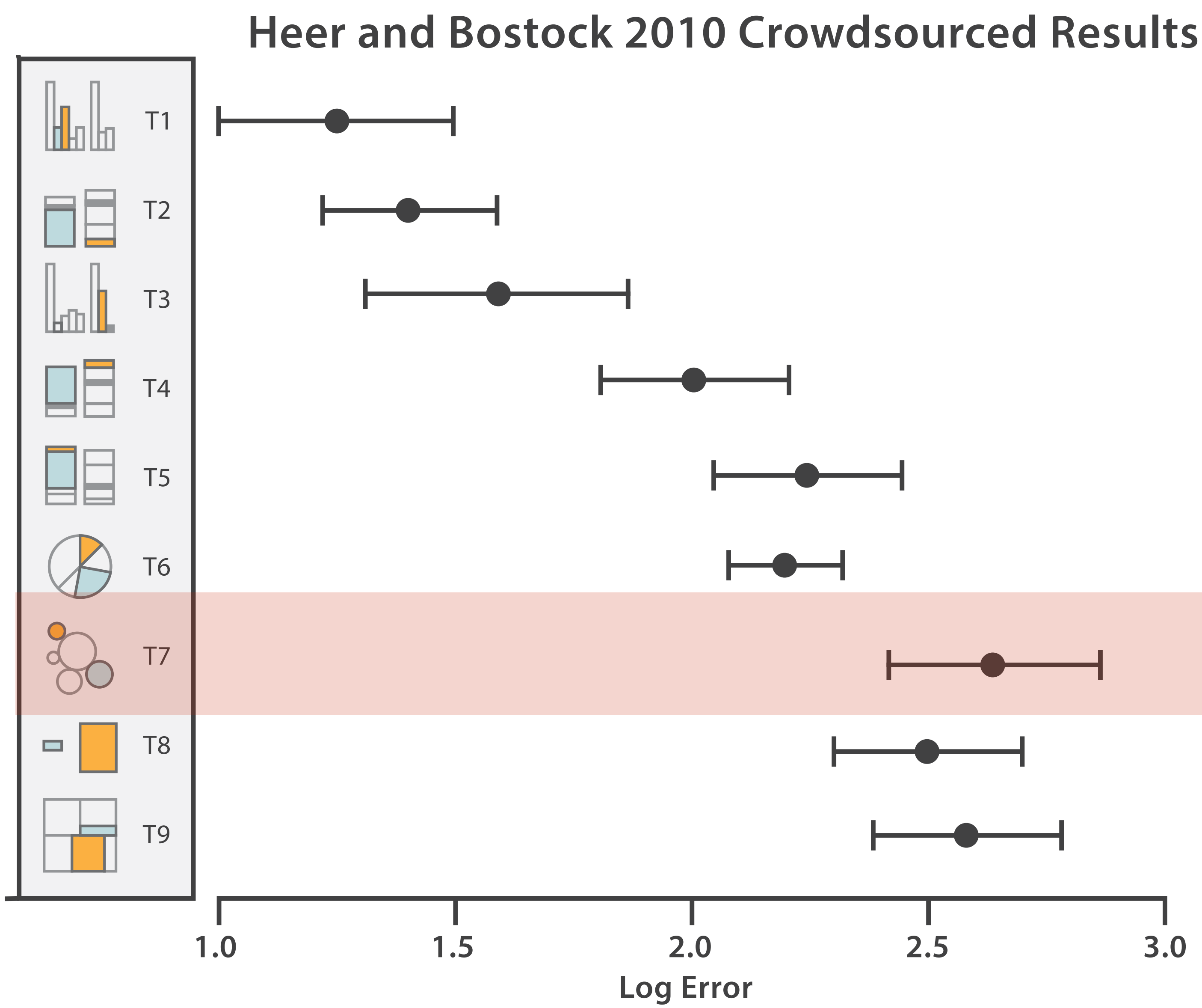




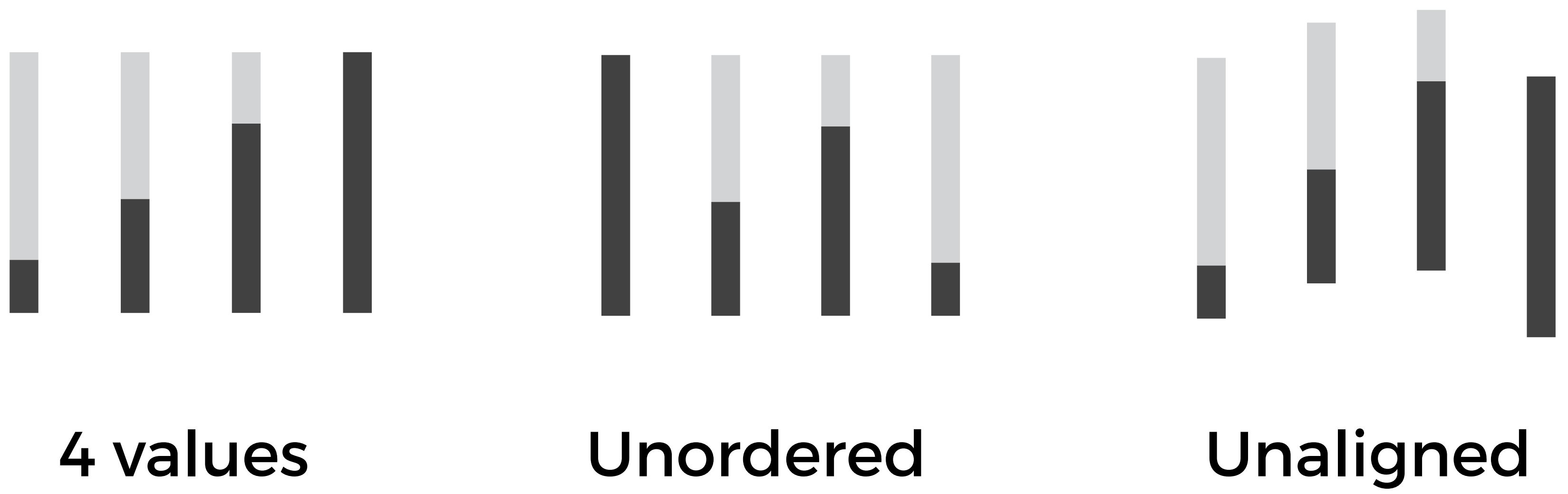
Relative Comparison

36px |

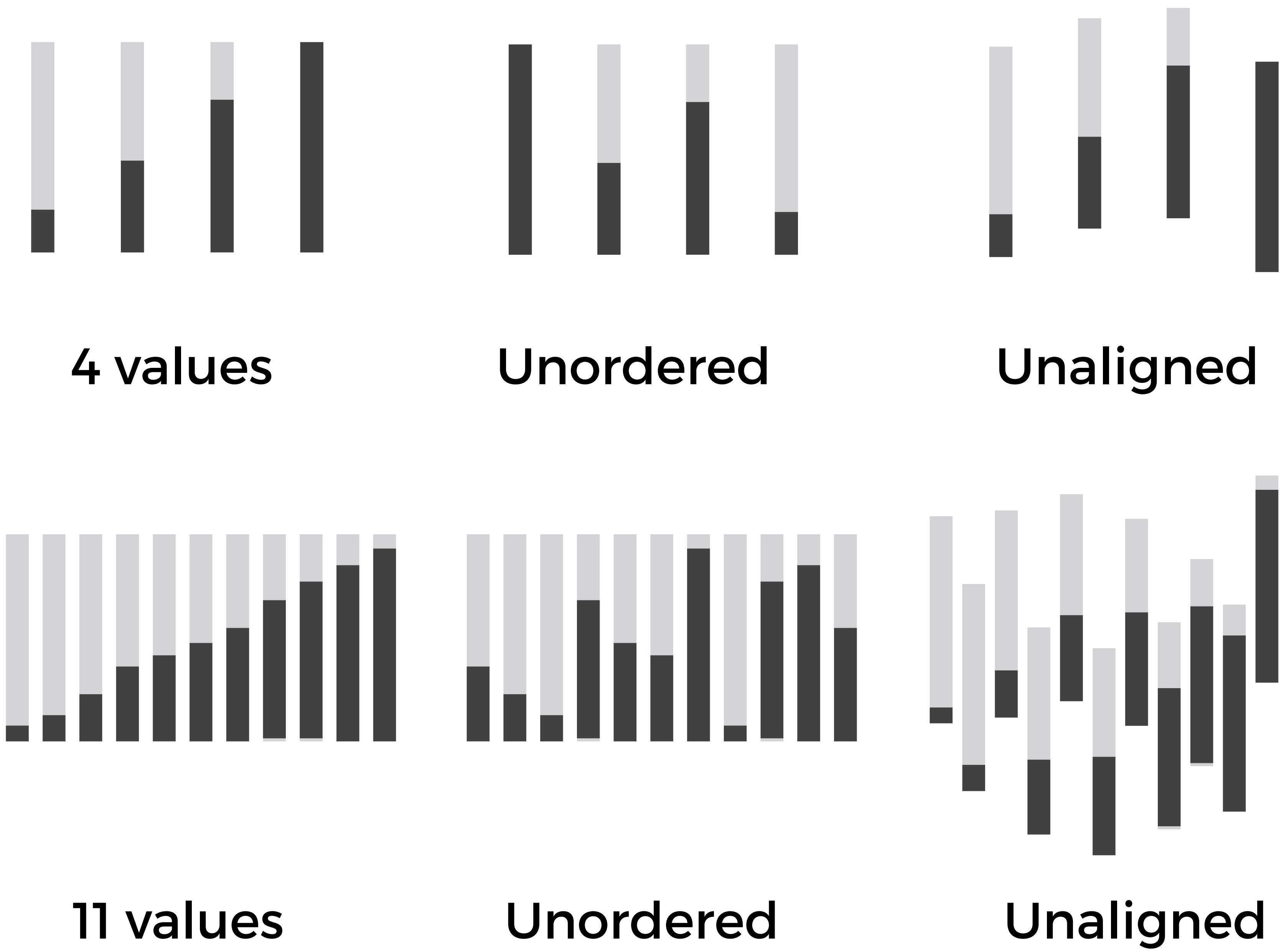


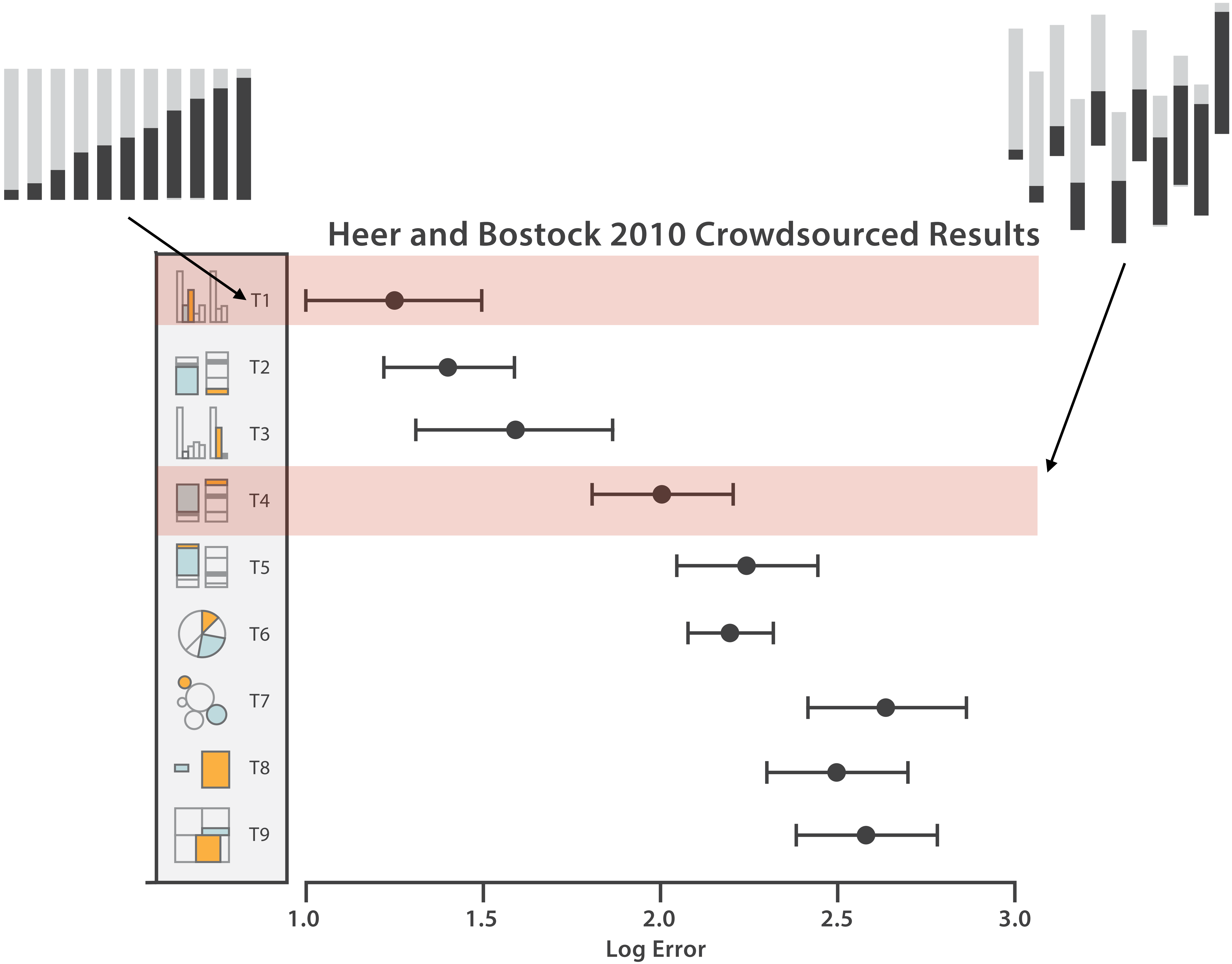


Relative Comparison



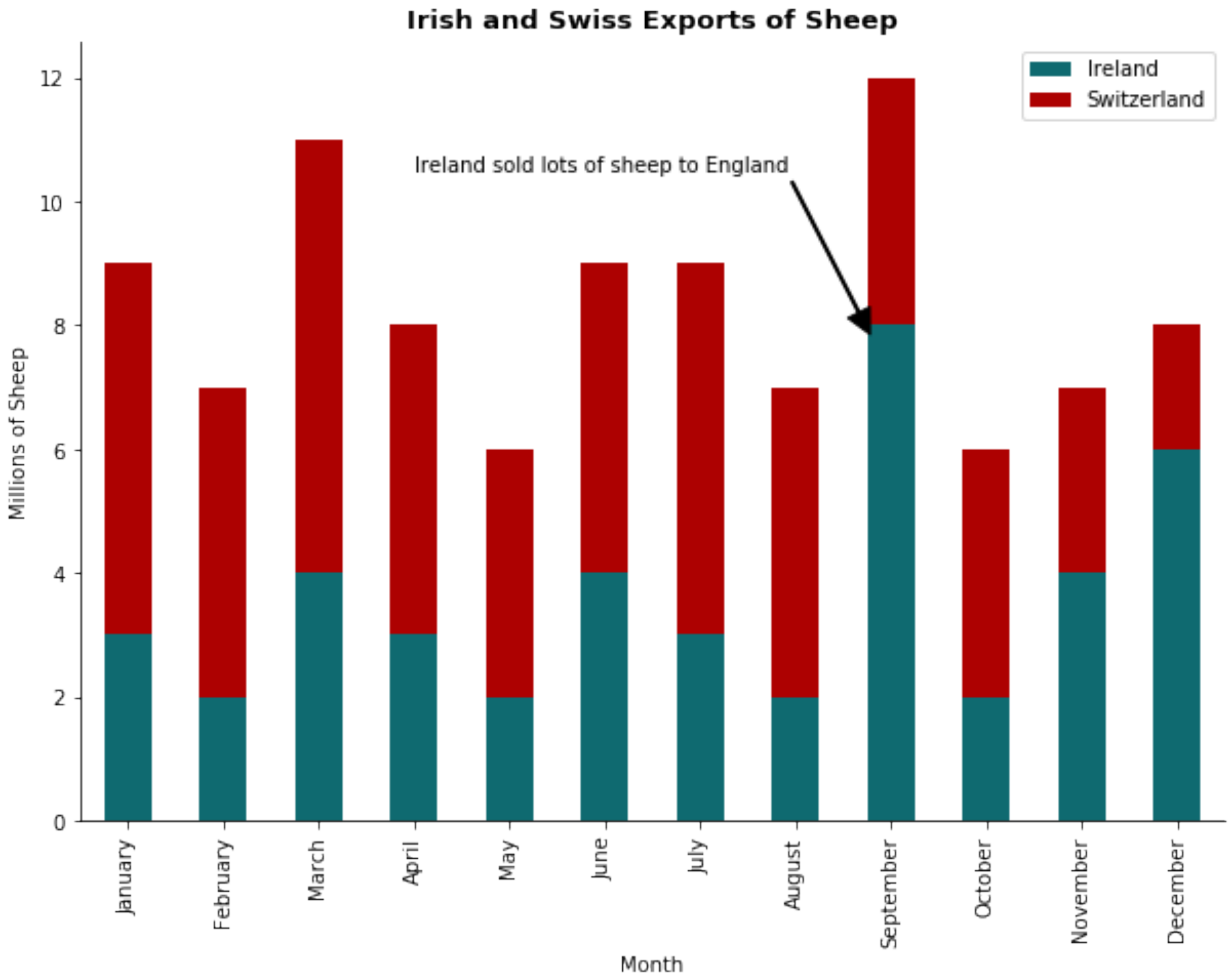
Relative Comparison





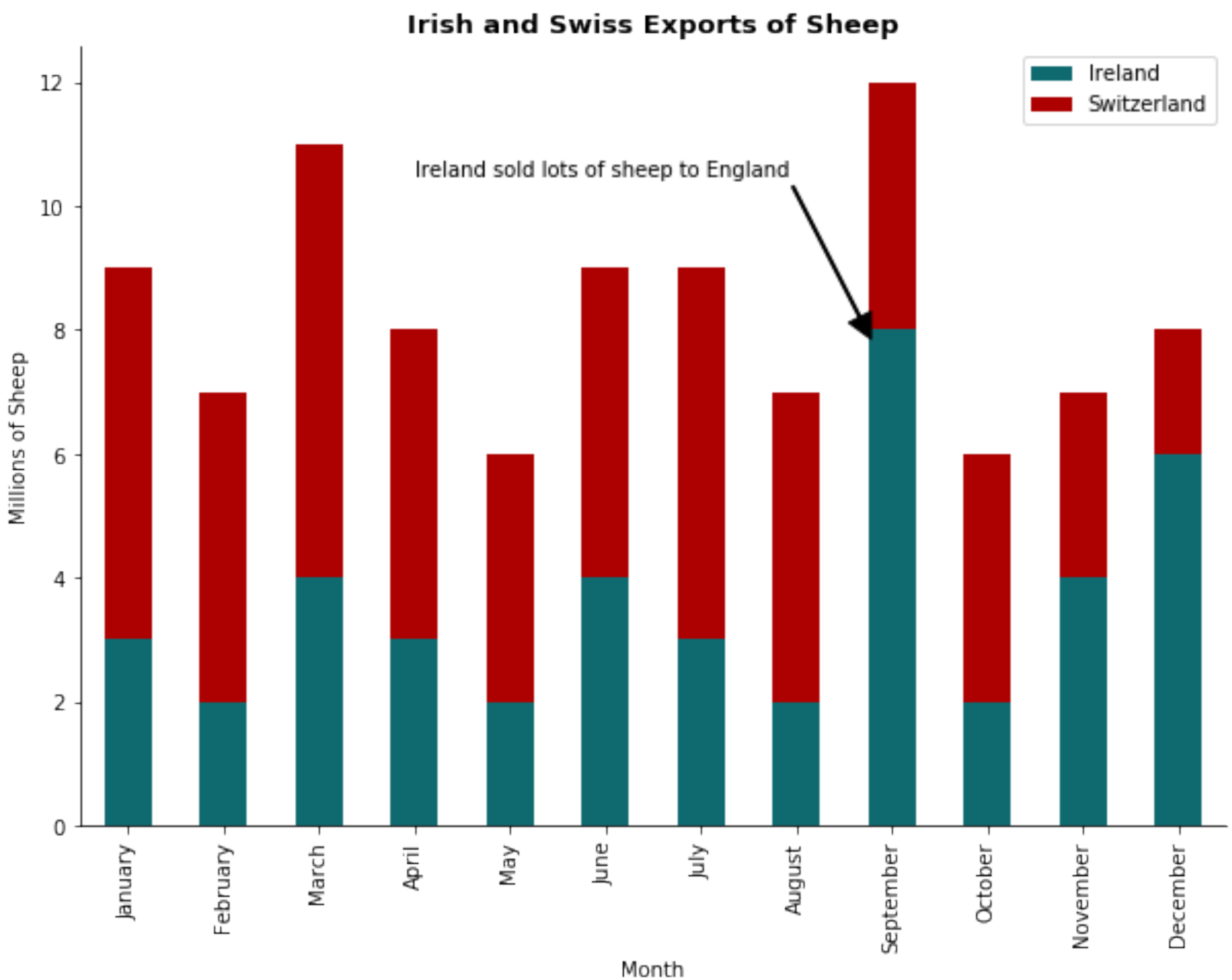
Heer and Bostock 2010 Crowdsourced Results
T1
T2
T3
T4
T5
T6
T7
T8
T9
1.0
1.5
2.0
2.5
3.0
Log Error

Relative Comparison

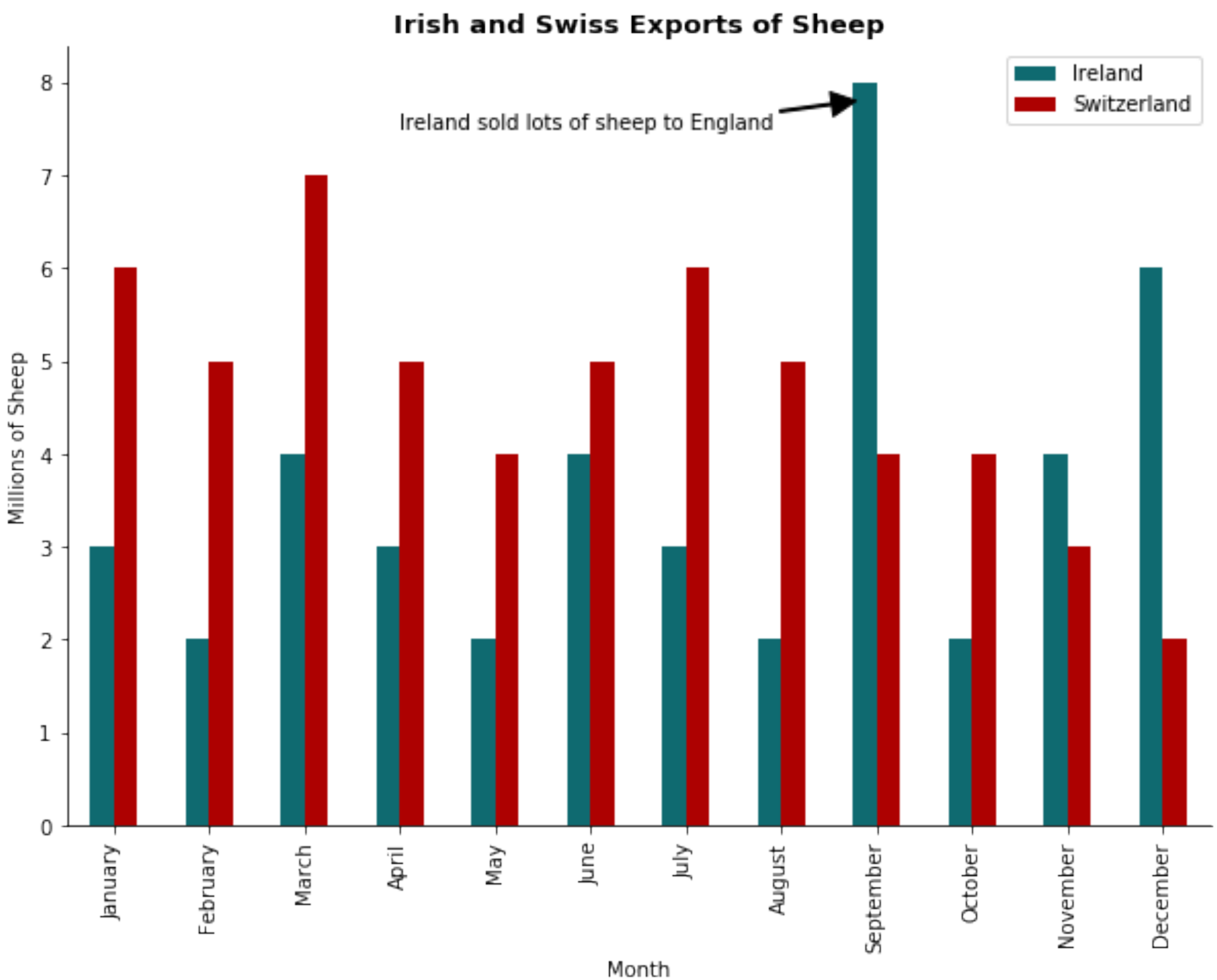


Relative Comparison

The problems with unaligned areas can be seen in stacked charts. A small number of values is ok, but too many and nothing will be interpretable.

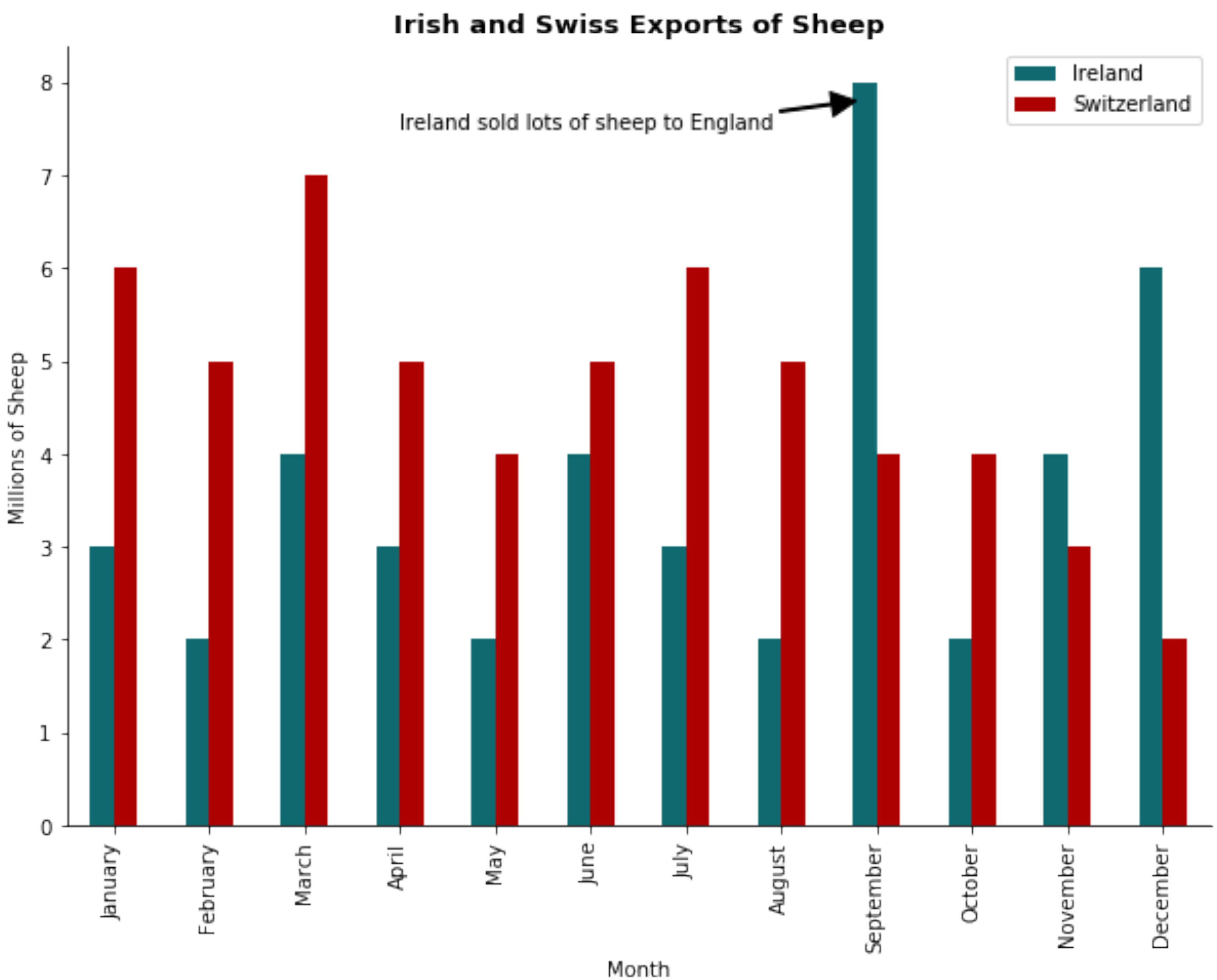


Relative Comparison

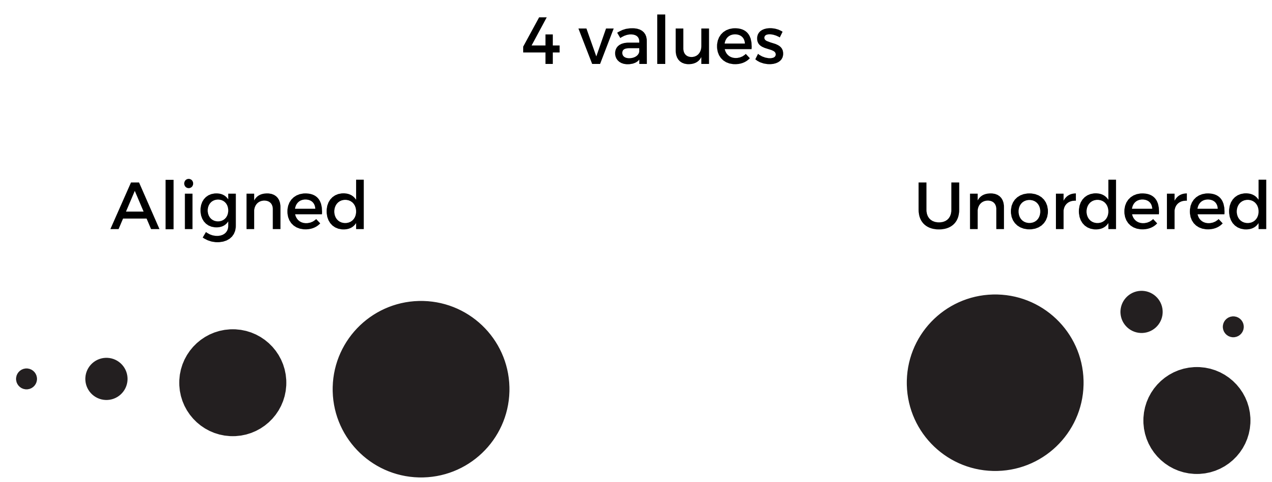


Relative Comparison

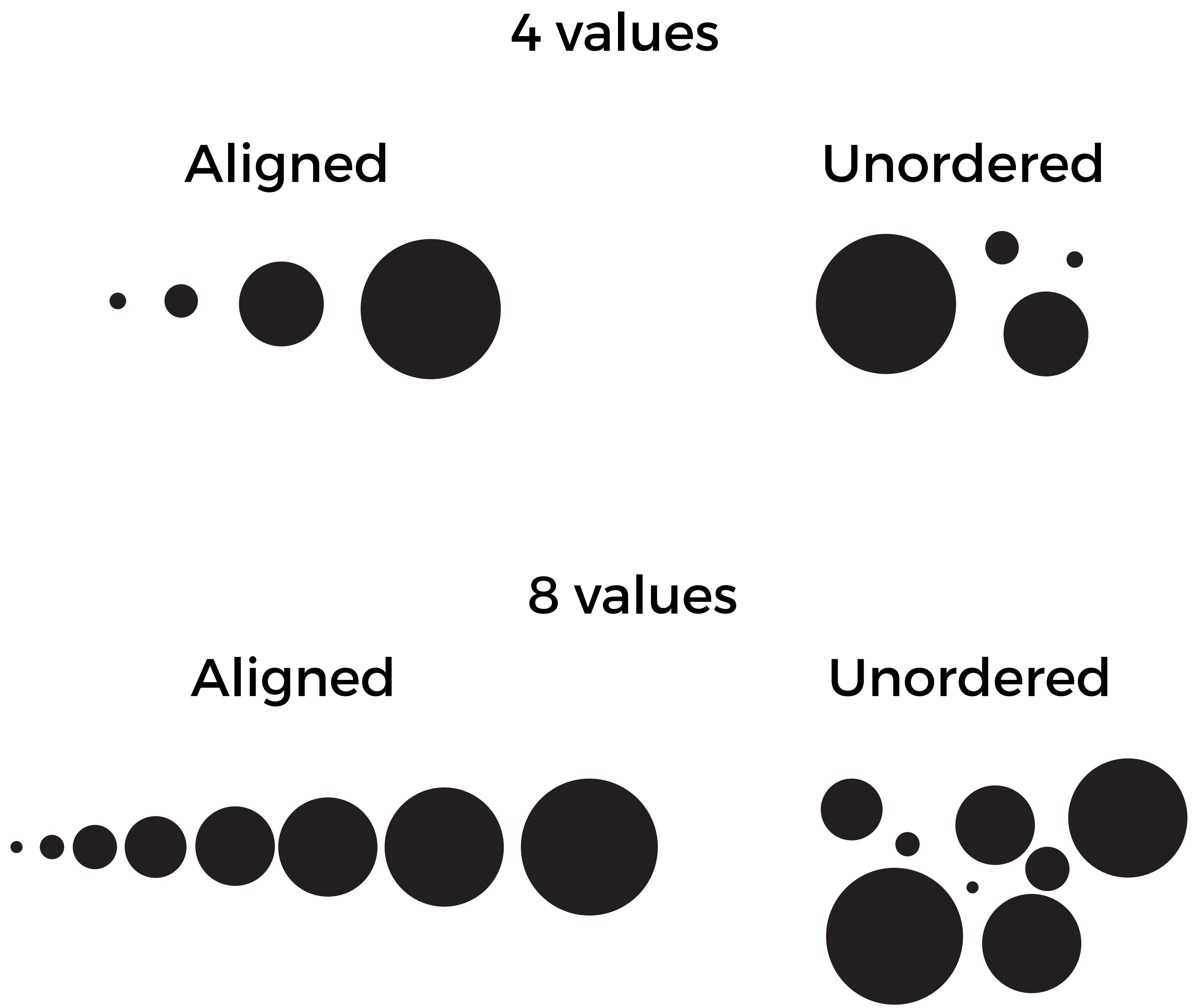
The problems with unaligned areas can be seen in stacked charts. A small number of values is ok, but too many and nothing will be interpretable.

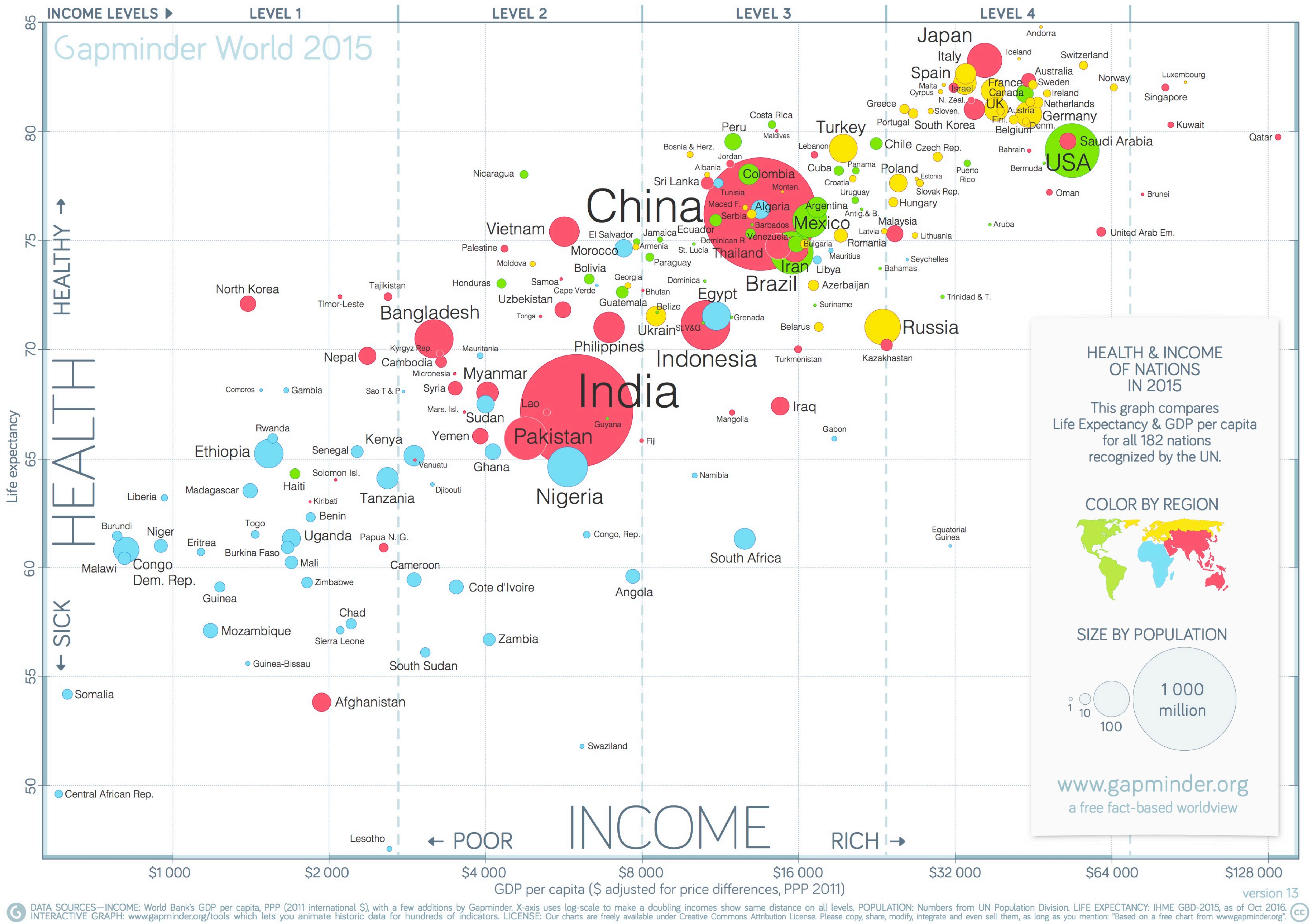


Relative Comparison

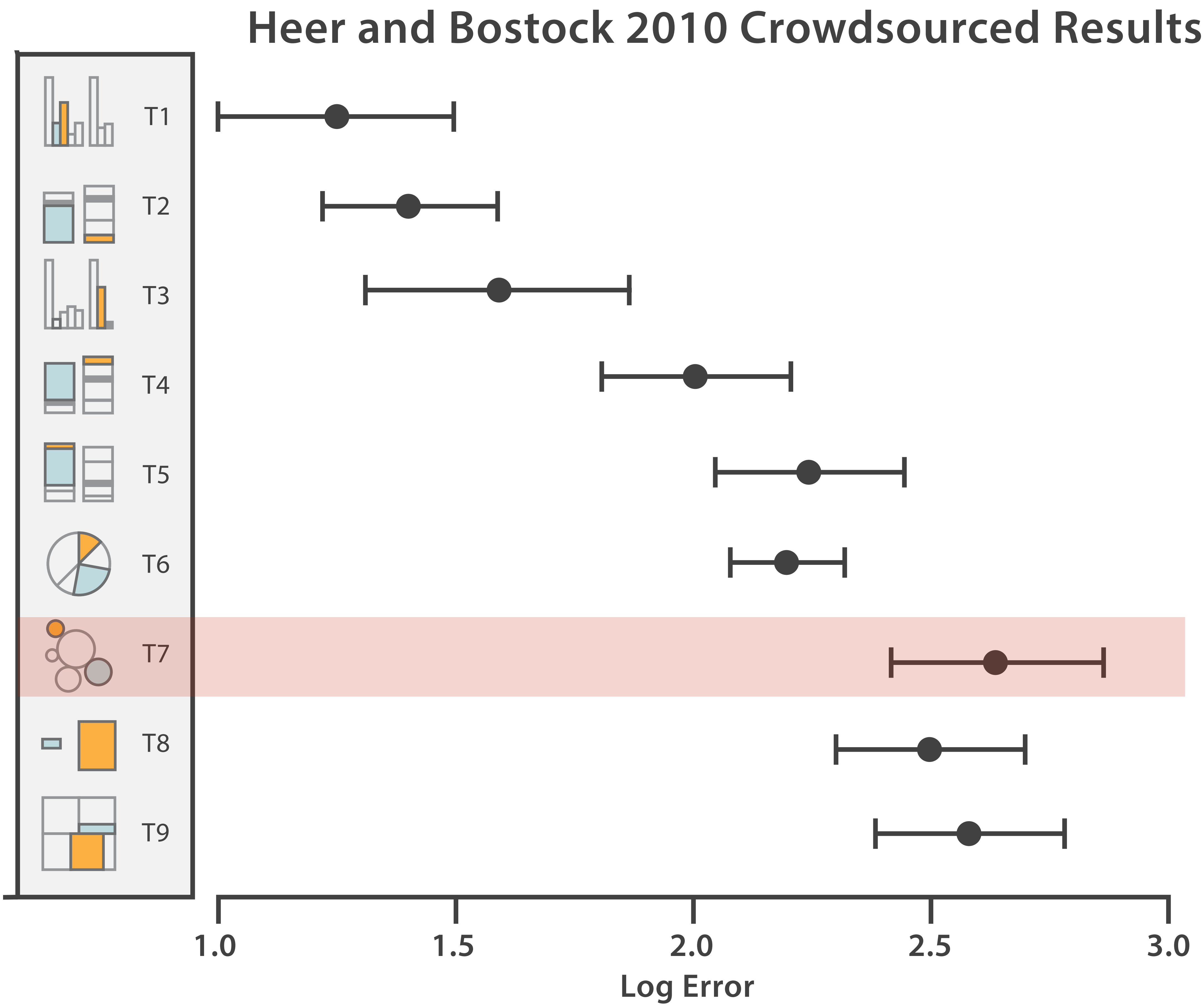


Relative Comparison





The infamous GAP minder chart is subject to such issues with relative comparison.



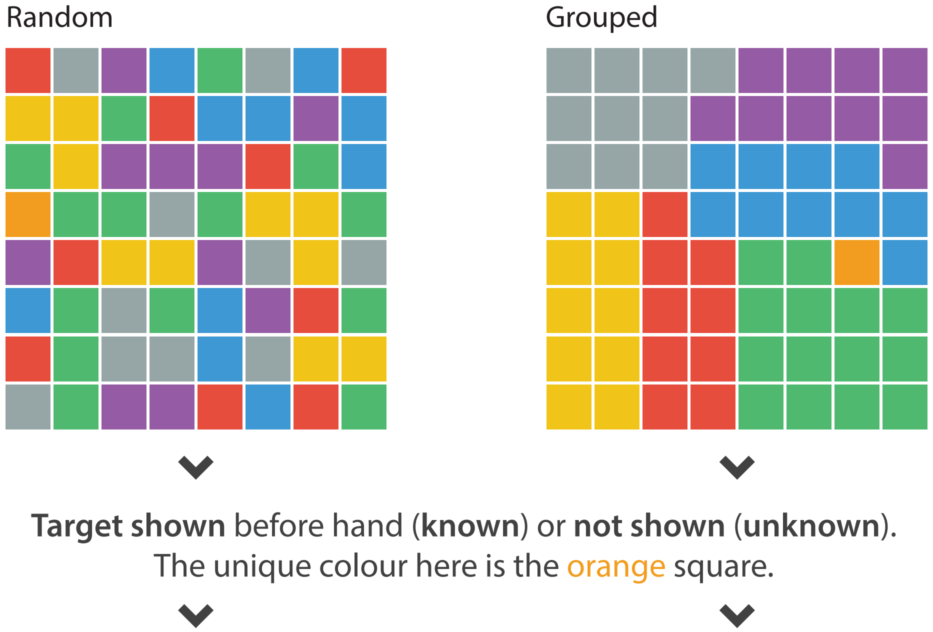
Heer and Bostock 2010 Crowdsourced Results
T1
T2
T3
T4
T5
T6
T7
T8
T9
1.0
1.5
2.0
2.5
3.0
Log Error

Relative Comparison

8 values

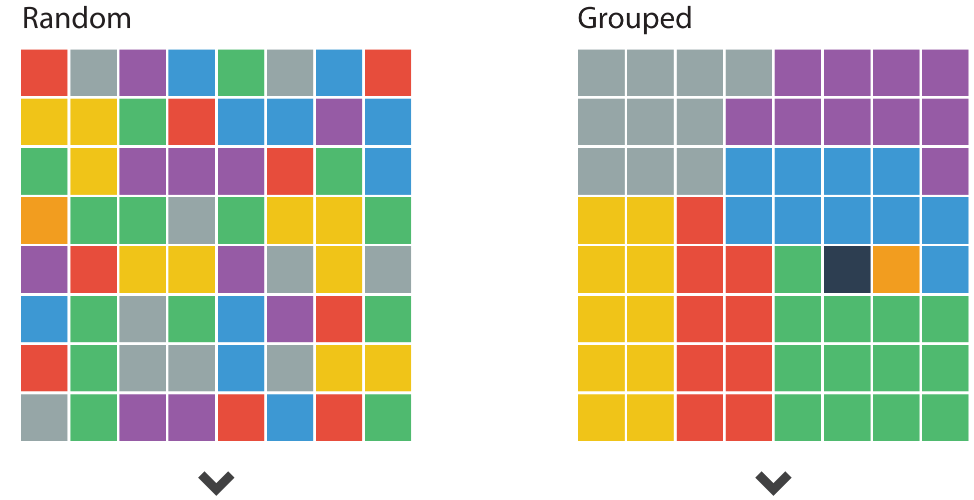
20 values

A) Known and Unknown Target Search



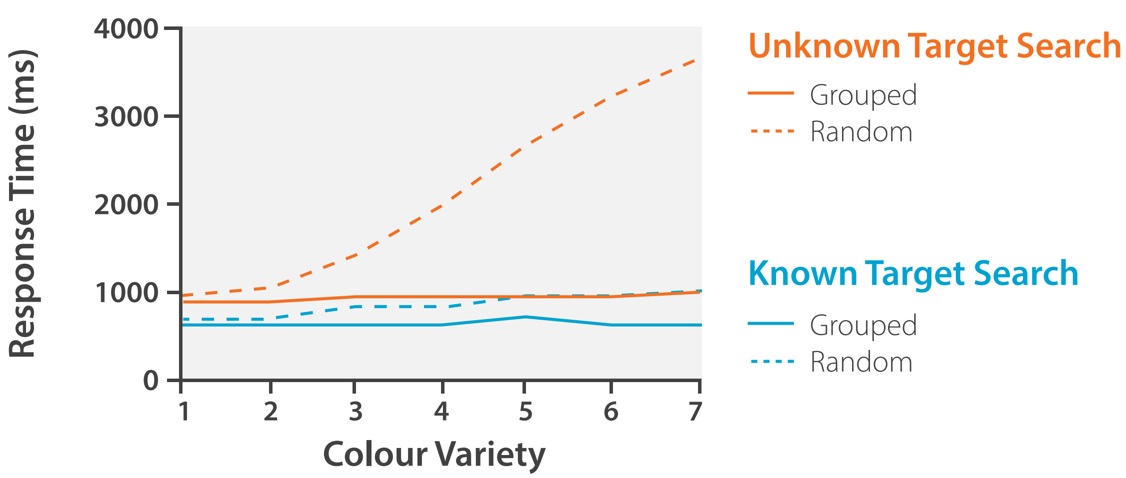
Target shown before hand (**known**) or **not shown** (**unknown**).
The unique colour here is the orange square.

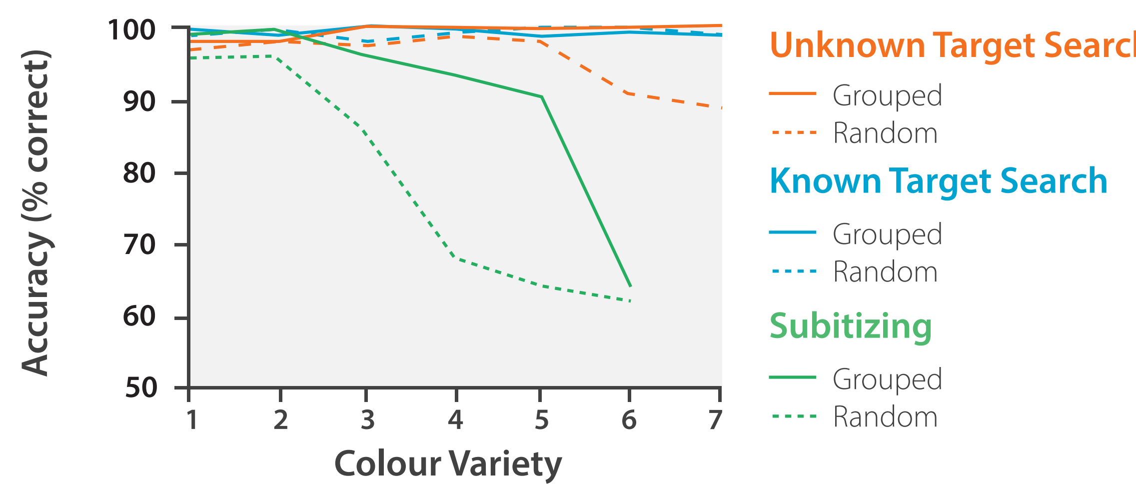
B) Subitizing (how many colours?)



Which grid has more colours?

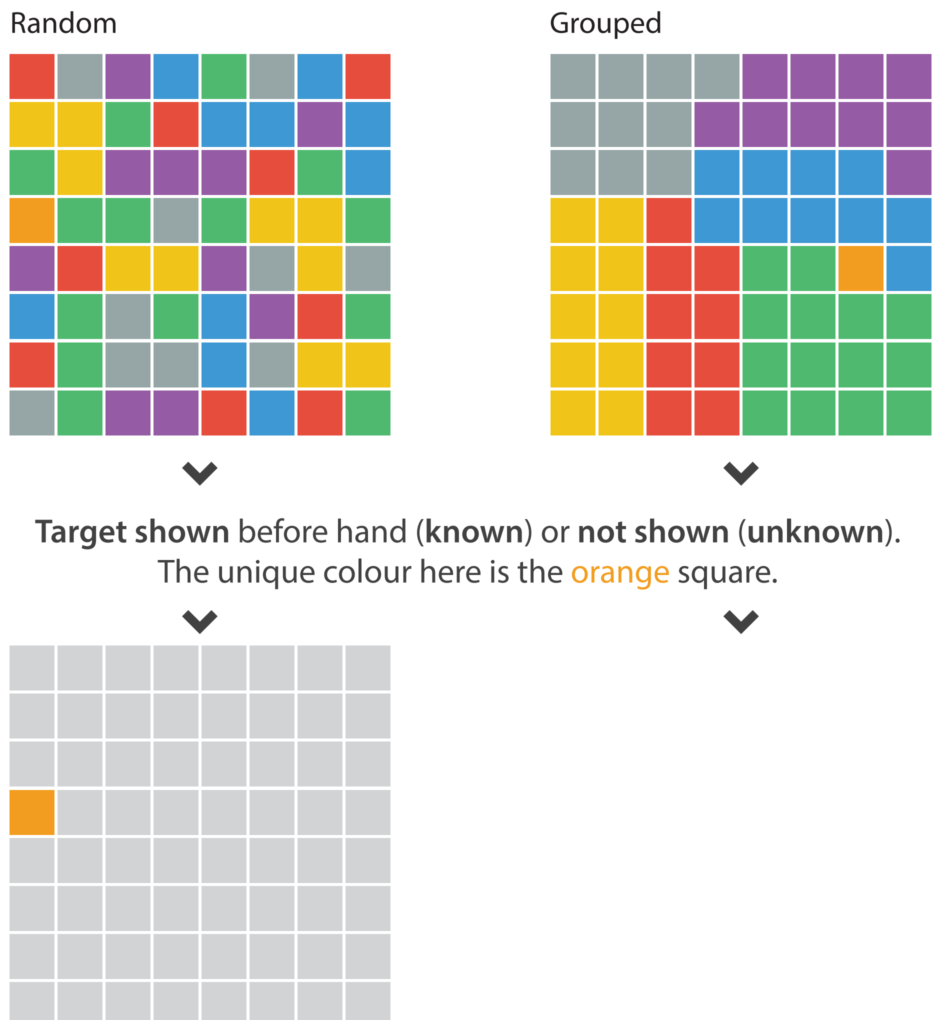
C) Response Time and Accuracy Results



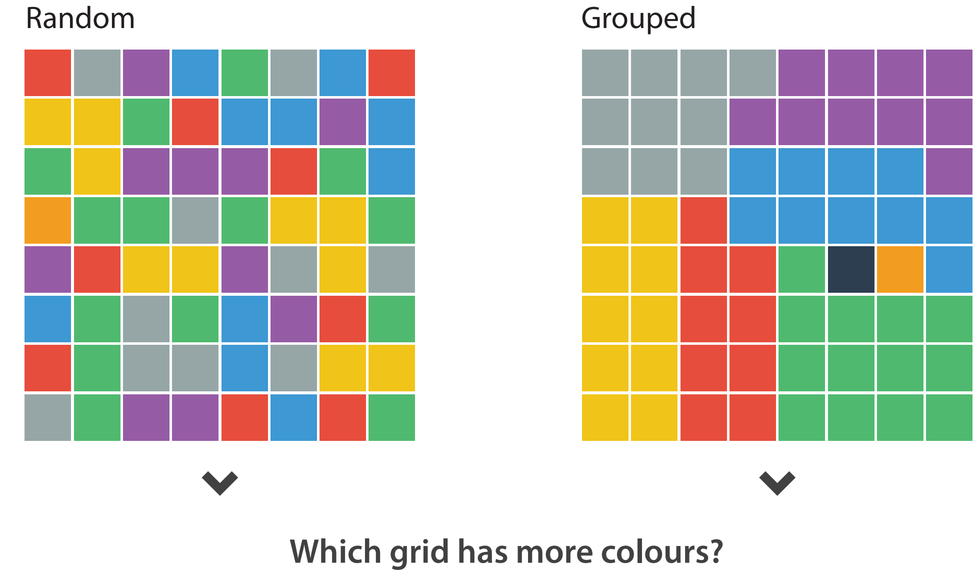


How Capacity Limits of Attention Influence Information Visualization Effectiveness.
Haroz S. and Whitney D., IEEE TVCG 2012

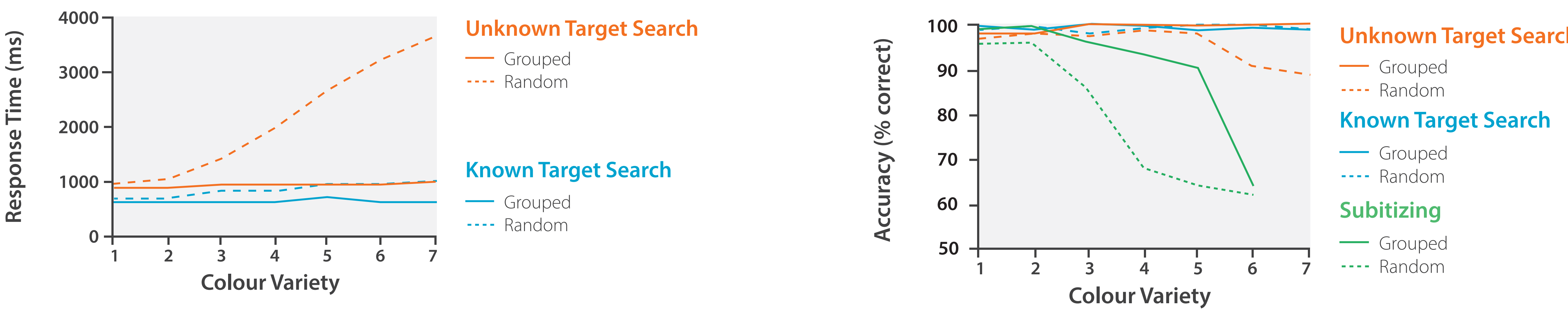
A) Known and Unknown Target Search



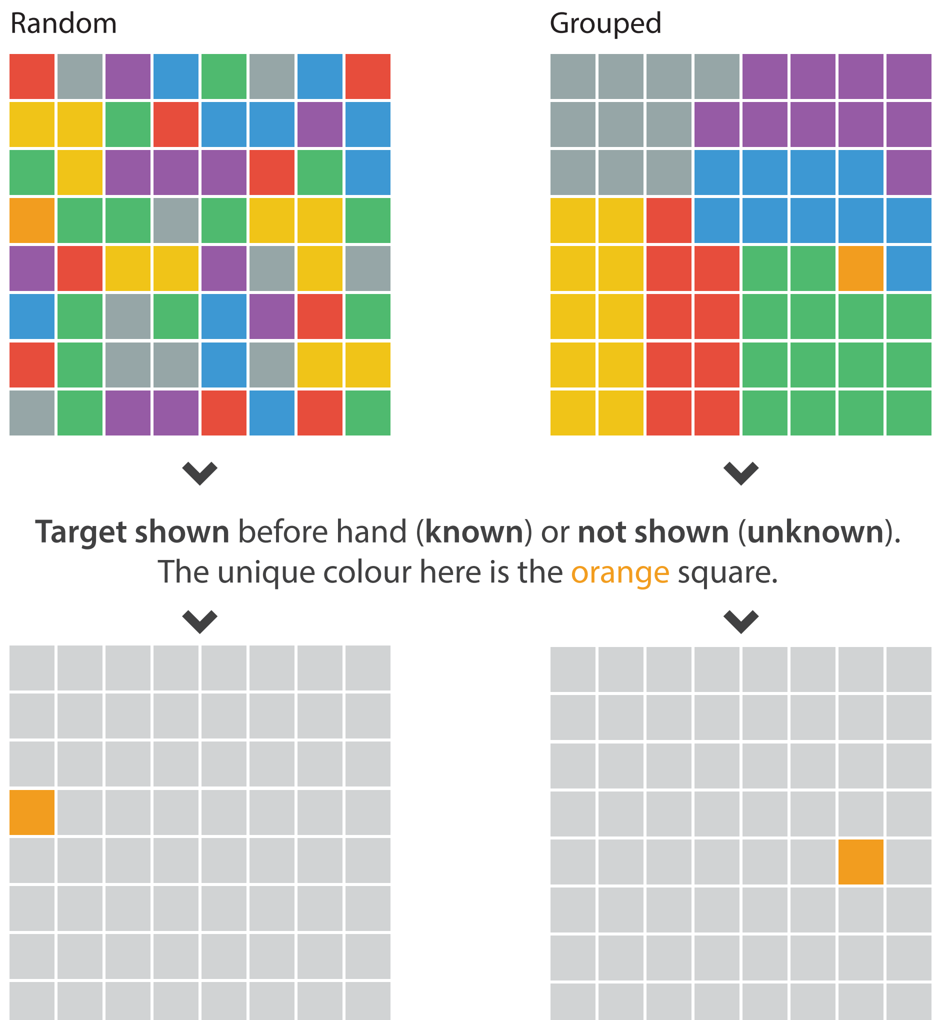
B) Subitizing (how many colours?)



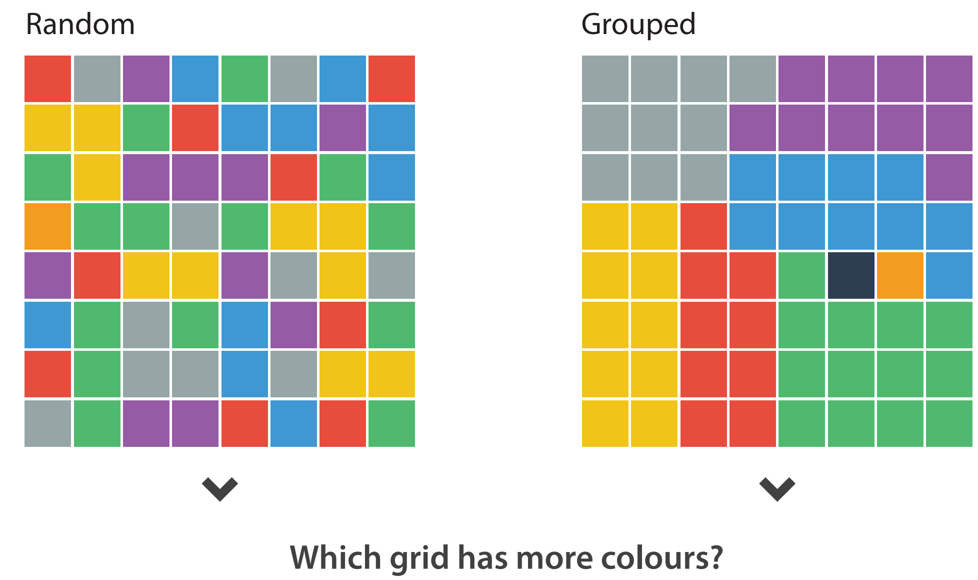
C) Response Time and Accuracy Results



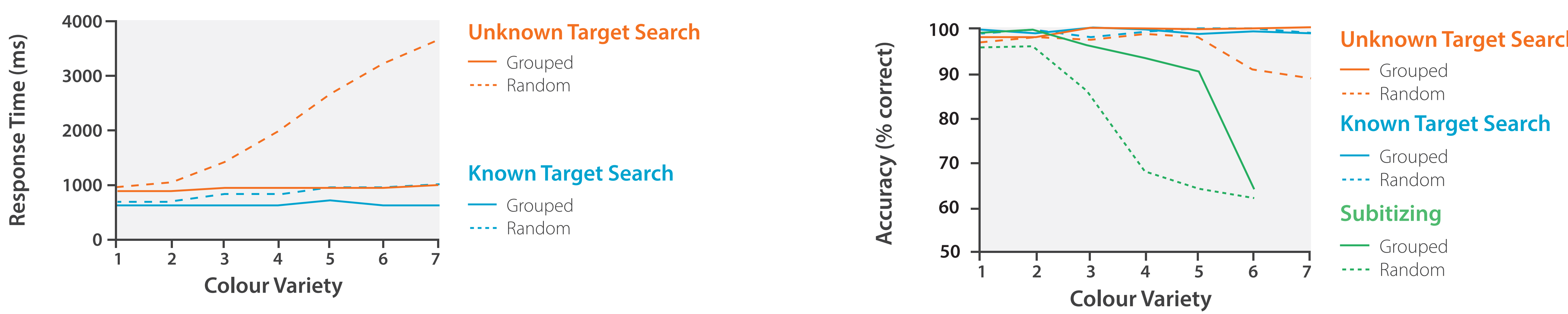
A) Known and Unknown Target Search



B) Subitizing (how many colours?)

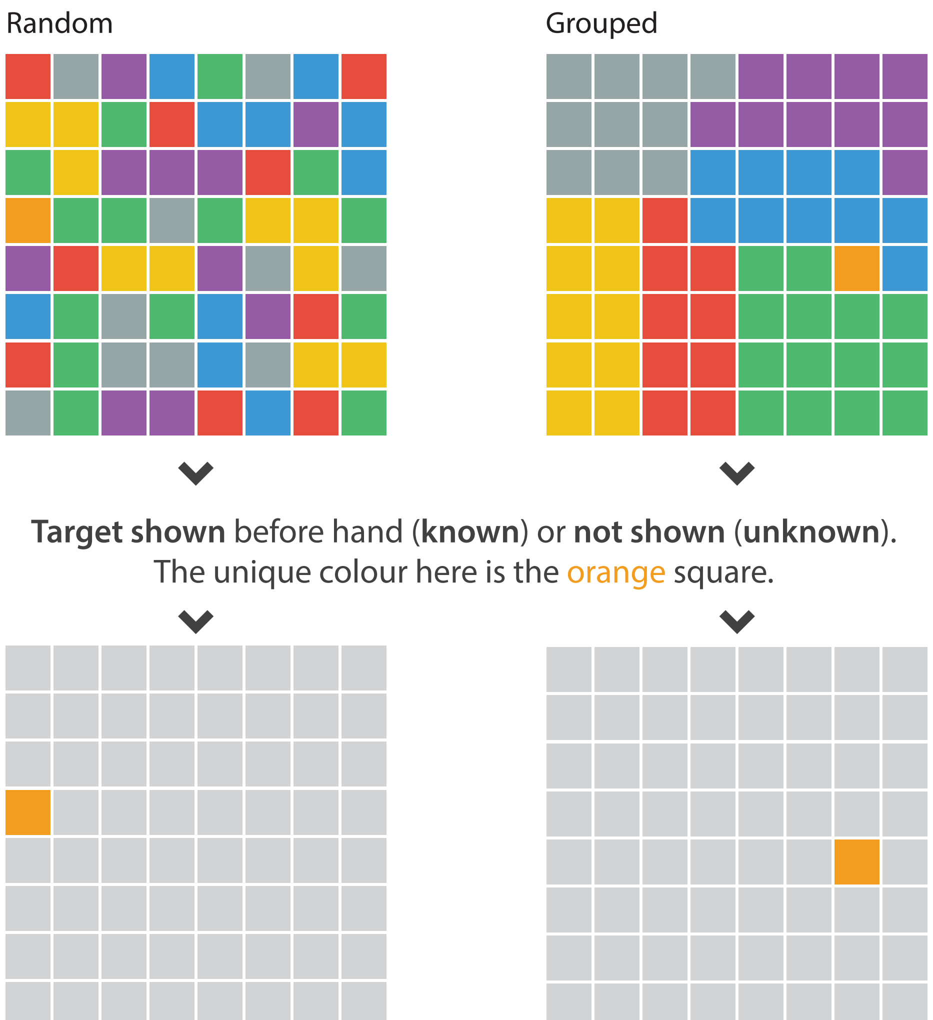


C) Response Time and Accuracy Results

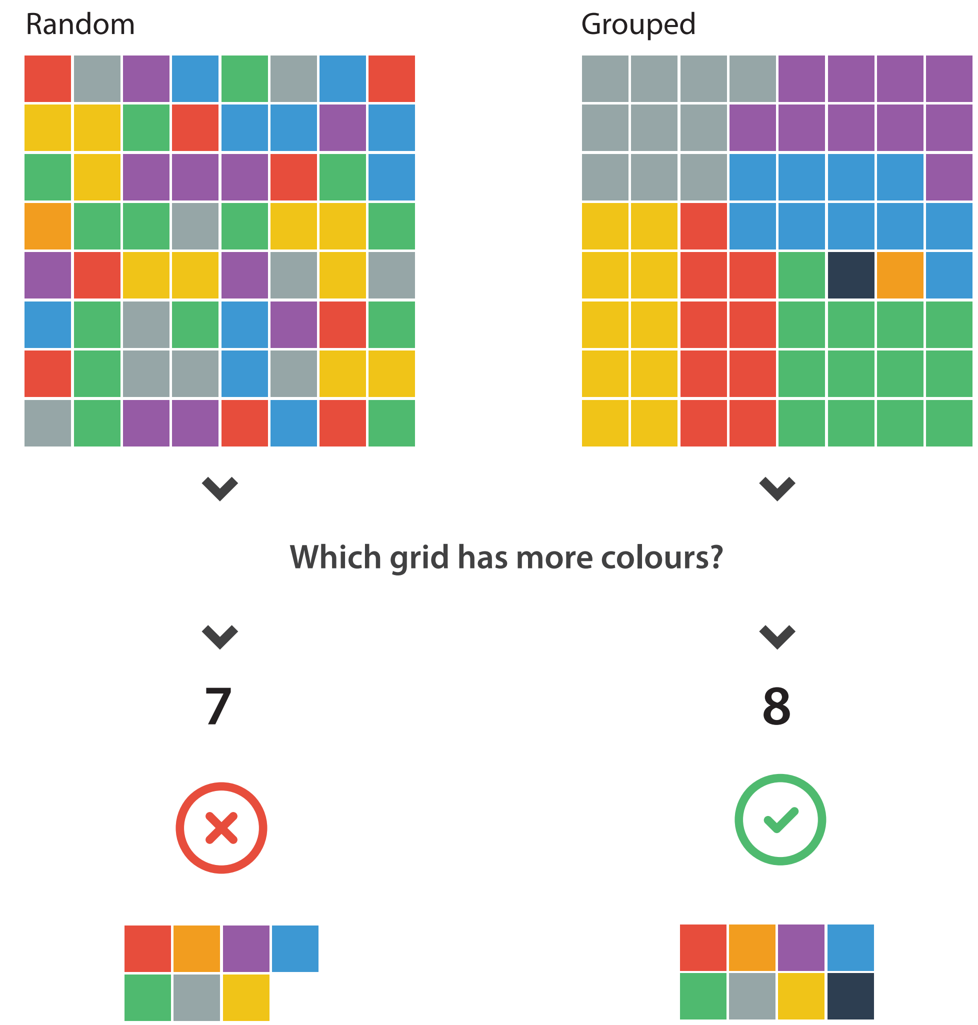


How Capacity Limits of Attention Influence Information Visualization Effectiveness.
Haroz S. and Whitney D., IEEE TVCG 2012

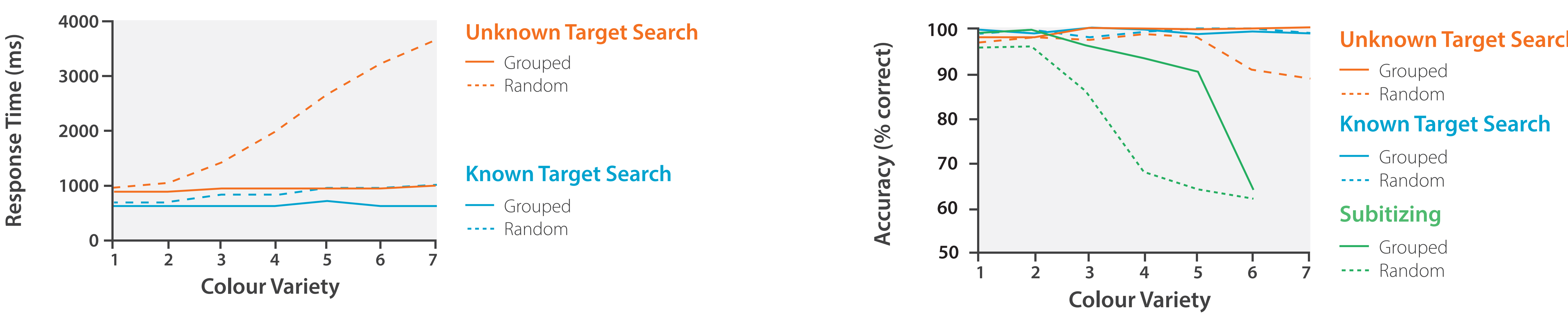
A) Known and Unknown Target Search



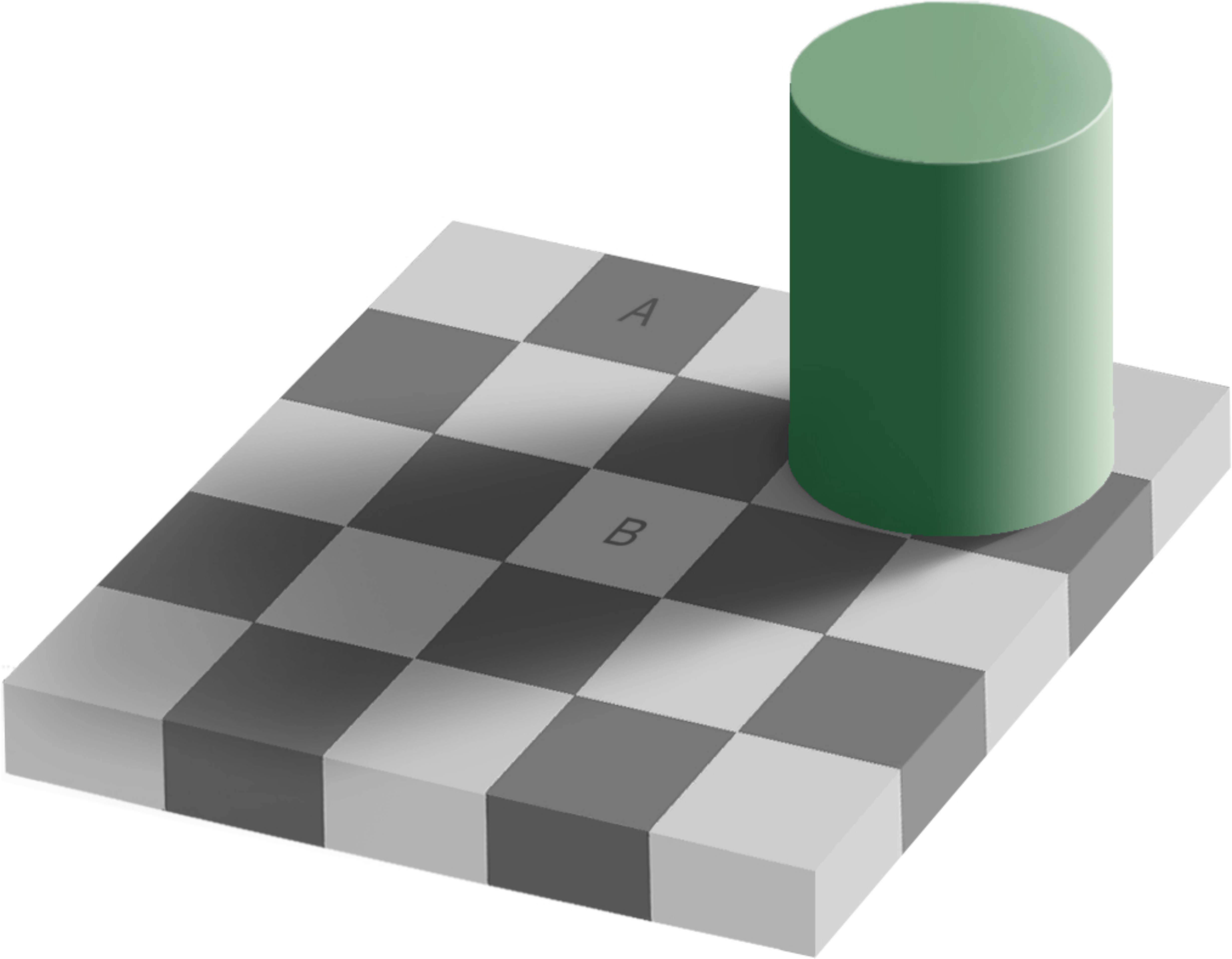
B) Subitizing (how many colours?)



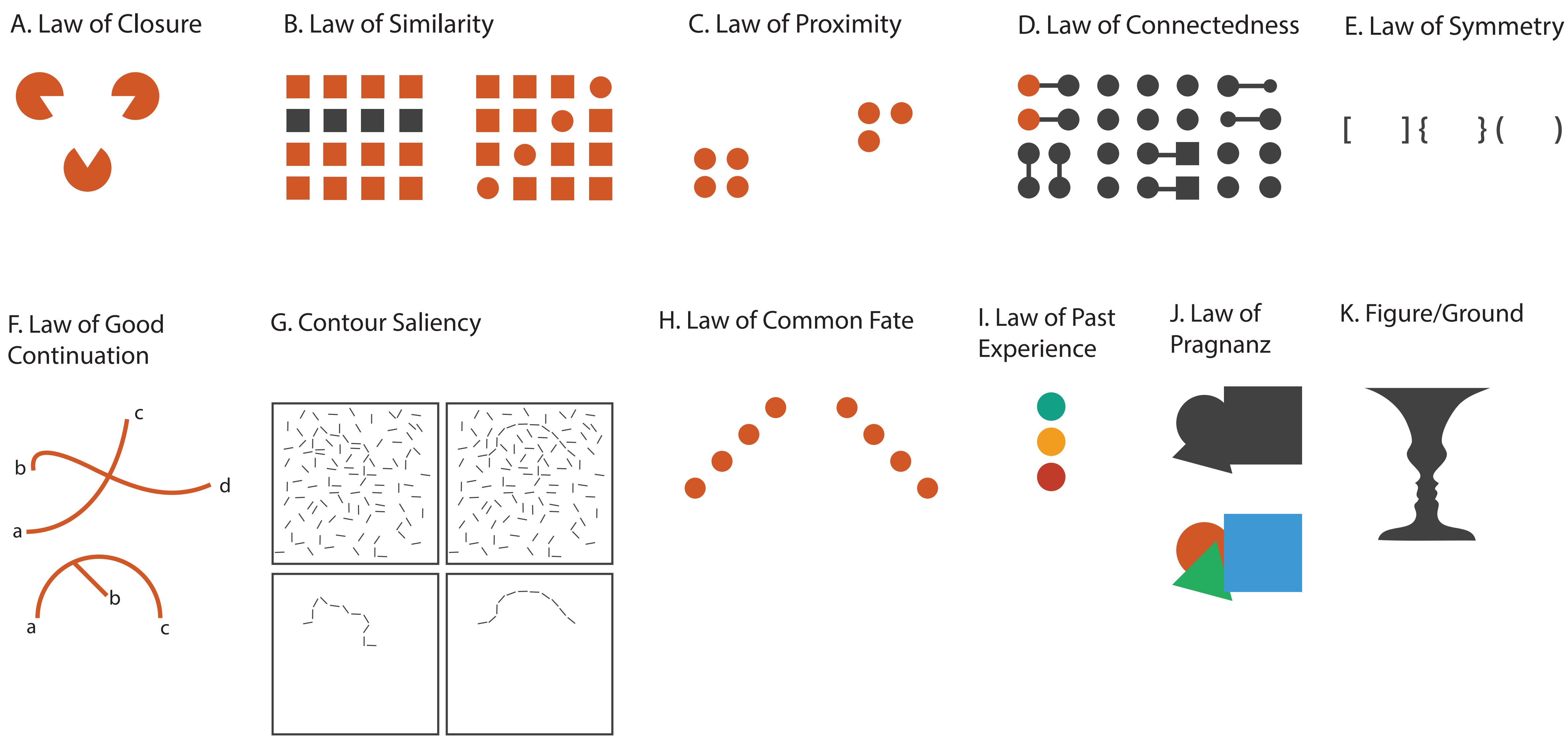
C) Response Time and Accuracy Results



How Capacity Limits of Attention Influence Information Visualization Effectiveness.
Haroz S. and Whitney D., IEEE TVCG 2012


A
B

Gestalt Laws



HOW

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

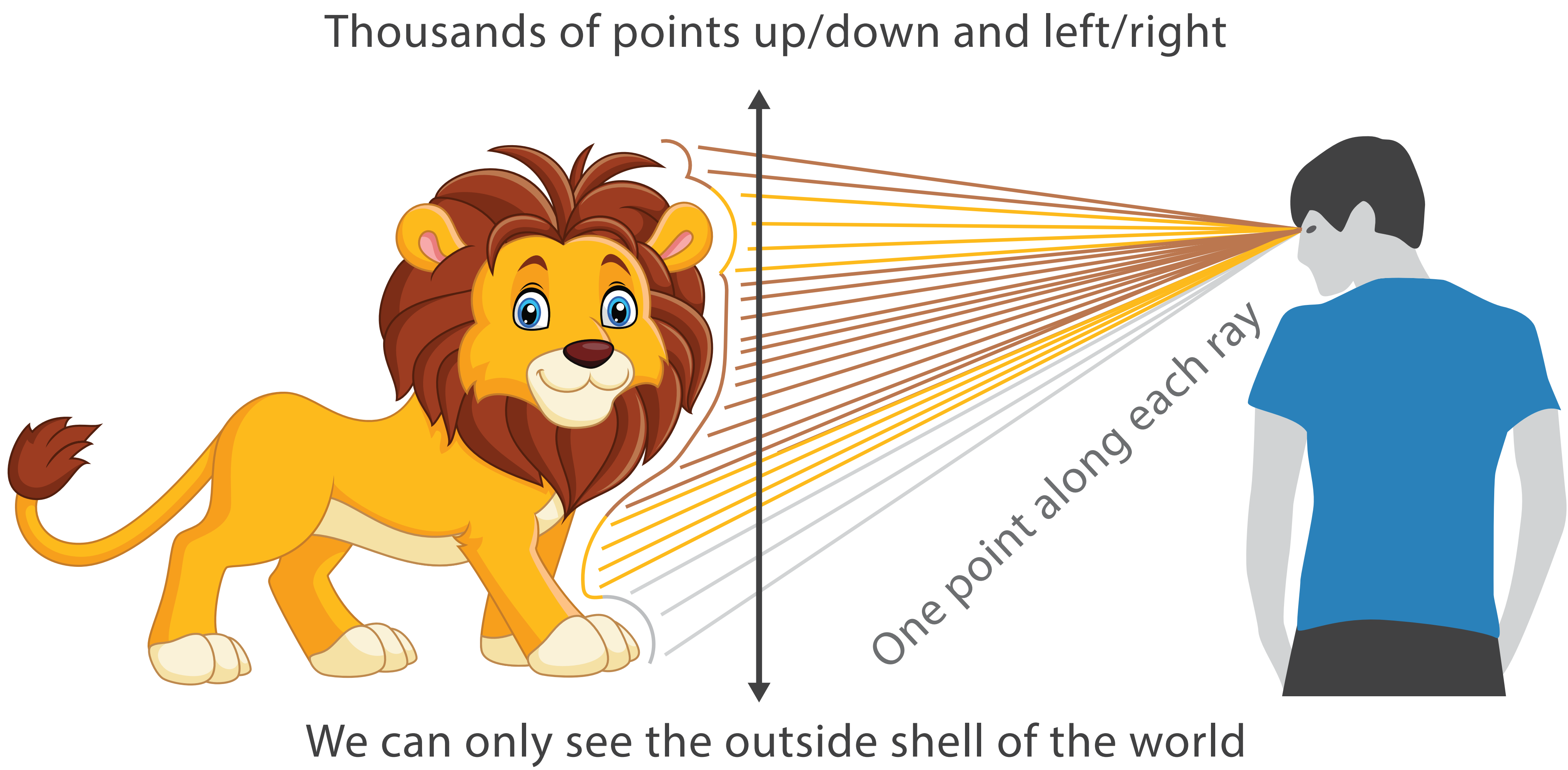
Some visual channels are more effective for some data types over others.

Some data has a **natural mapping** that our brains expect given certain types of data

There are many visual tricks that can be observed due to how the visual system works

We don't see in 3D, and we have difficulties interpreting information on the Z-axis.

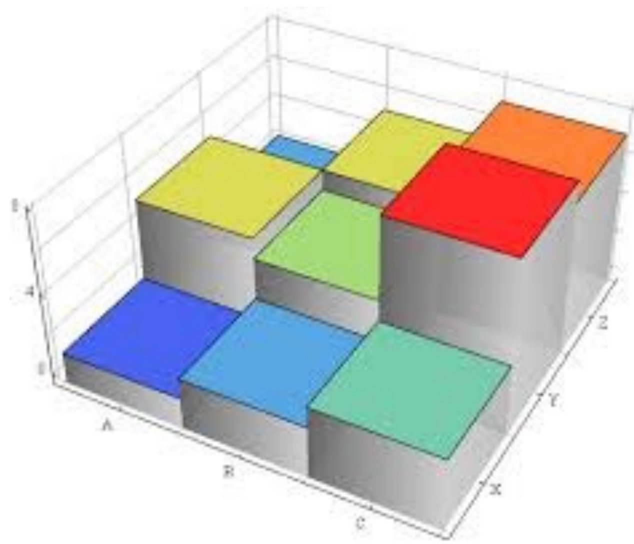
2D always wins...

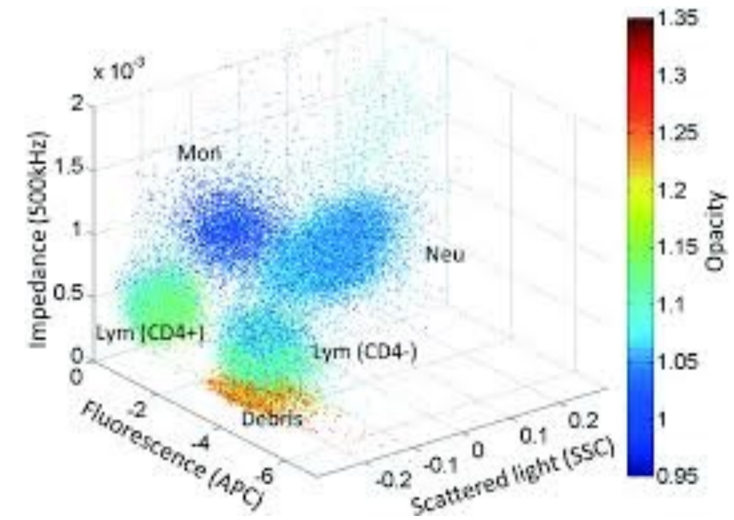


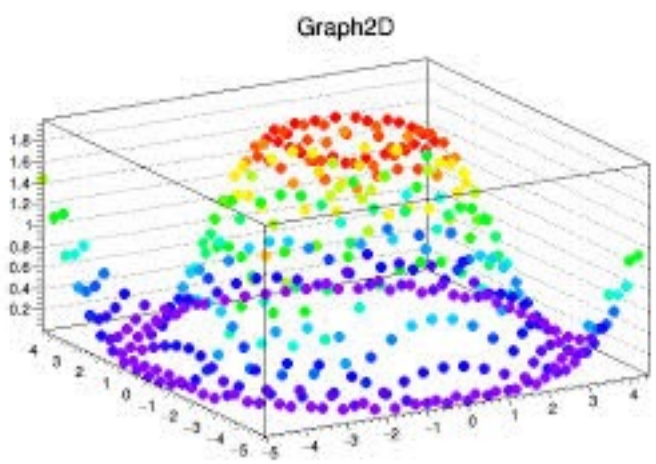
Our visual system is not good at interpreting information on the z-axis.

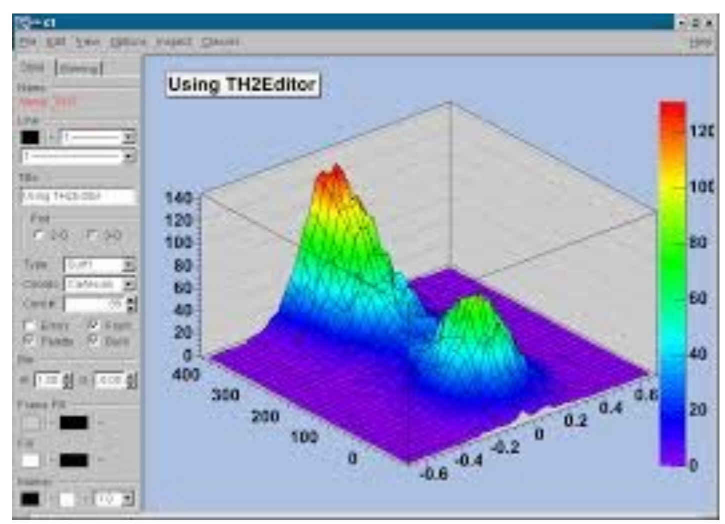
*3D is normally only used for exploration of inherently 3D information, such as medical imaging data…

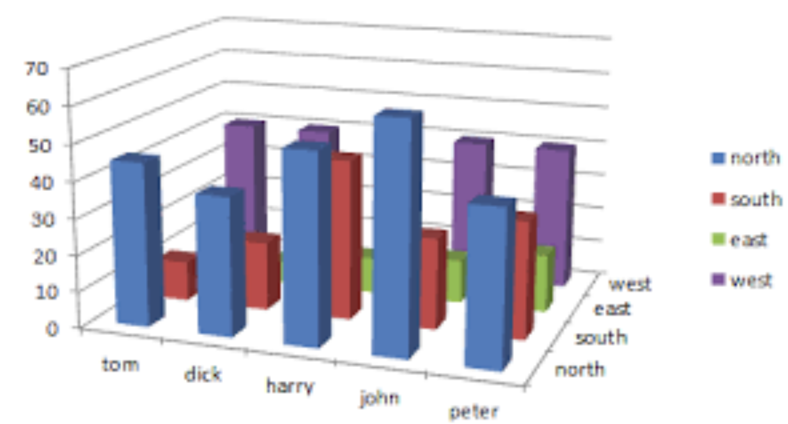
2D always wins...





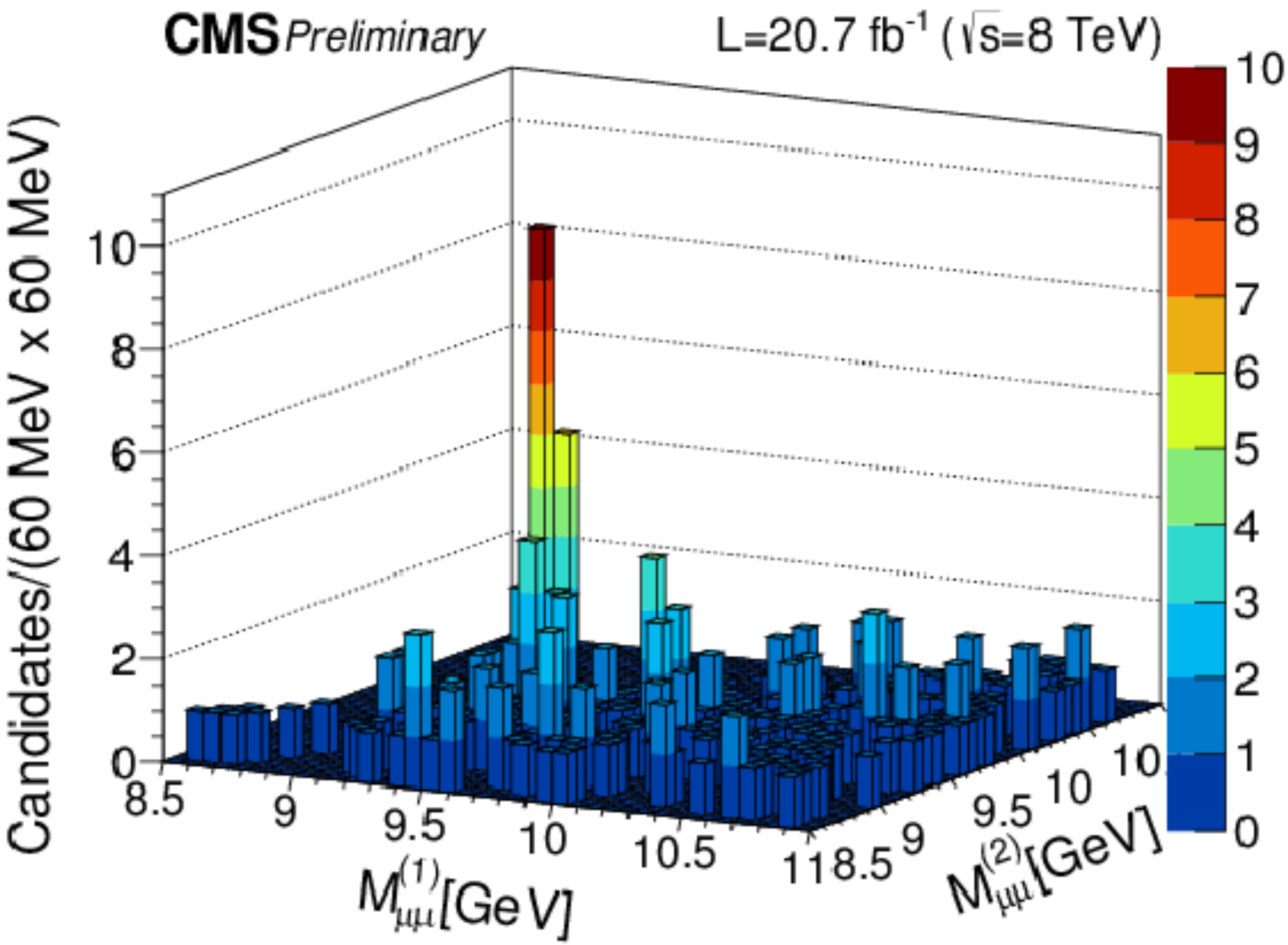






These options, taken randomly from google image searches so how widely 3D is abused in information visualisation. All of these charts are manipulating our perception of the data by using the Z axis to occlude information…it would be avoided in 2D.

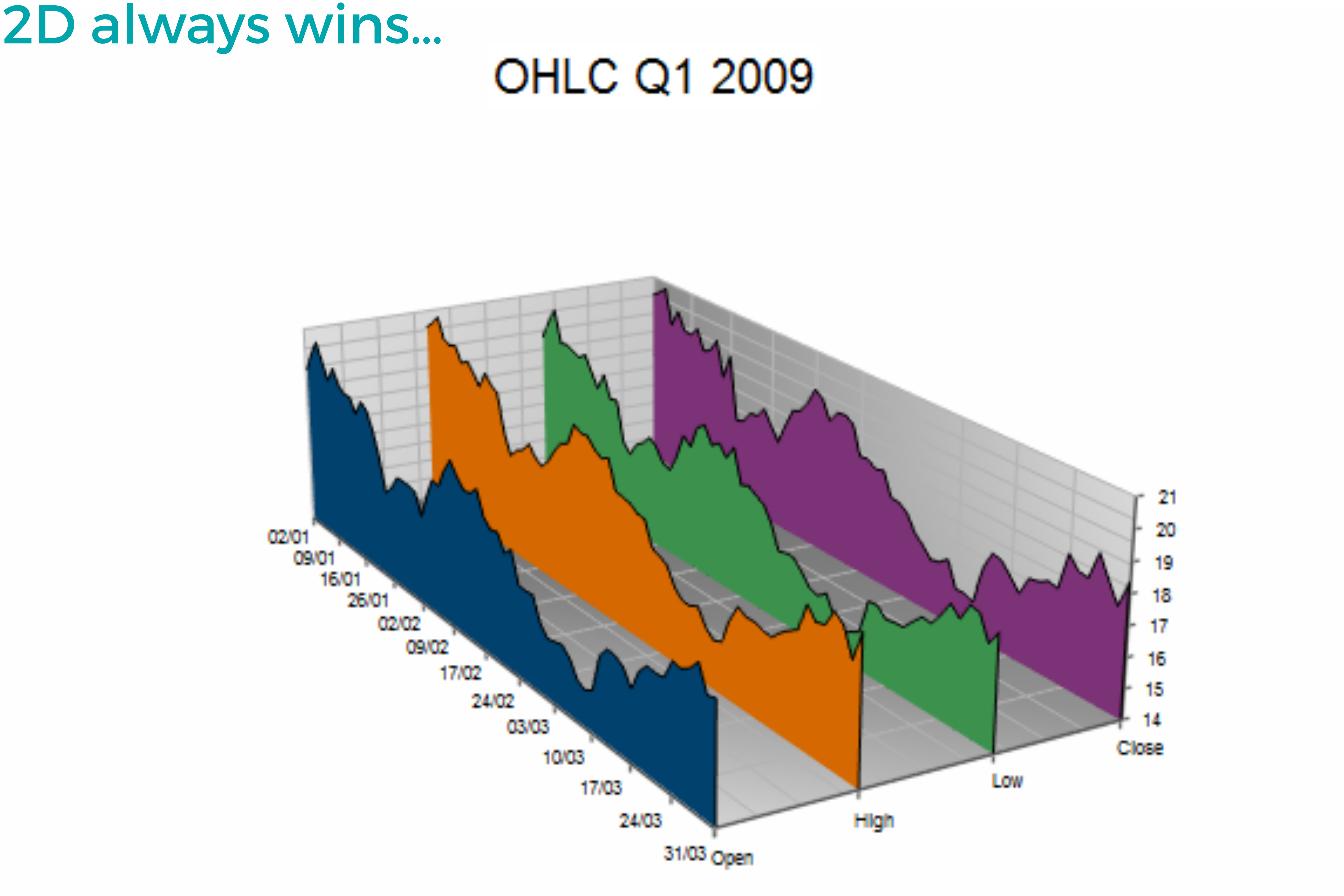
2D always wins...



3D hides information. Is there anything behind the large bars? We'll never know.

http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/BPH-14-008/index.html

2D always wins...



3D is totally useless in this example. It only makes the nearest points look bigger, and the further away points smaller than they are.

[Image from https://www.teraplot.com/financial](https://www.teraplot.com/financial)

HOW

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

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

Some data has a **natural mapping** that our brains expect given certain types of data

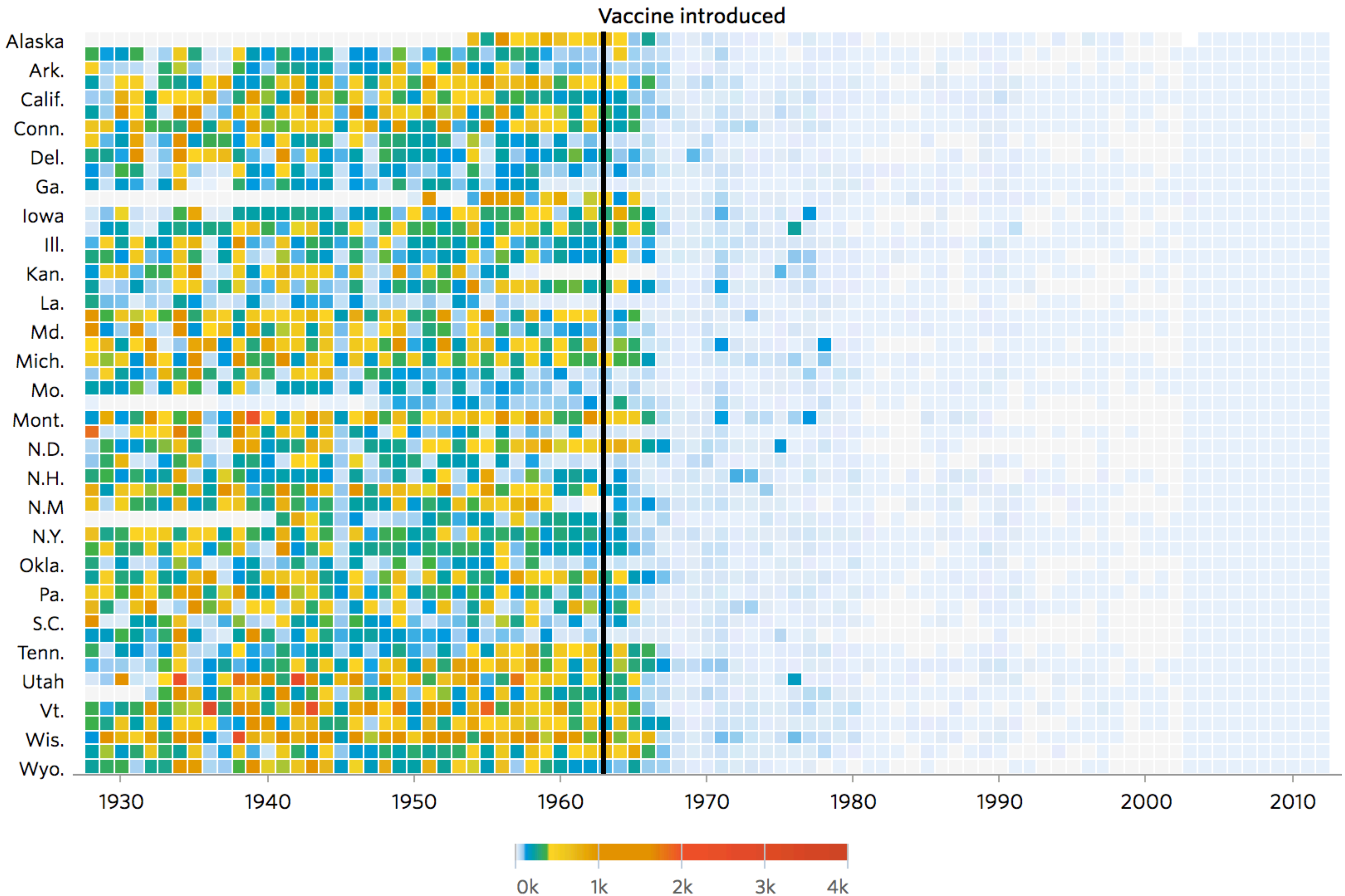
There are many visual tricks that can be observed due to how the visual system works

We don’t see in 3D, and we have difficulties interpreting information on the Z-axis.

Colour

Colour

Measles



http://graphics.wsj.com/infectious-diseases-and-vaccines/

Colour

The simplest, yet most abused of all visual encodings.

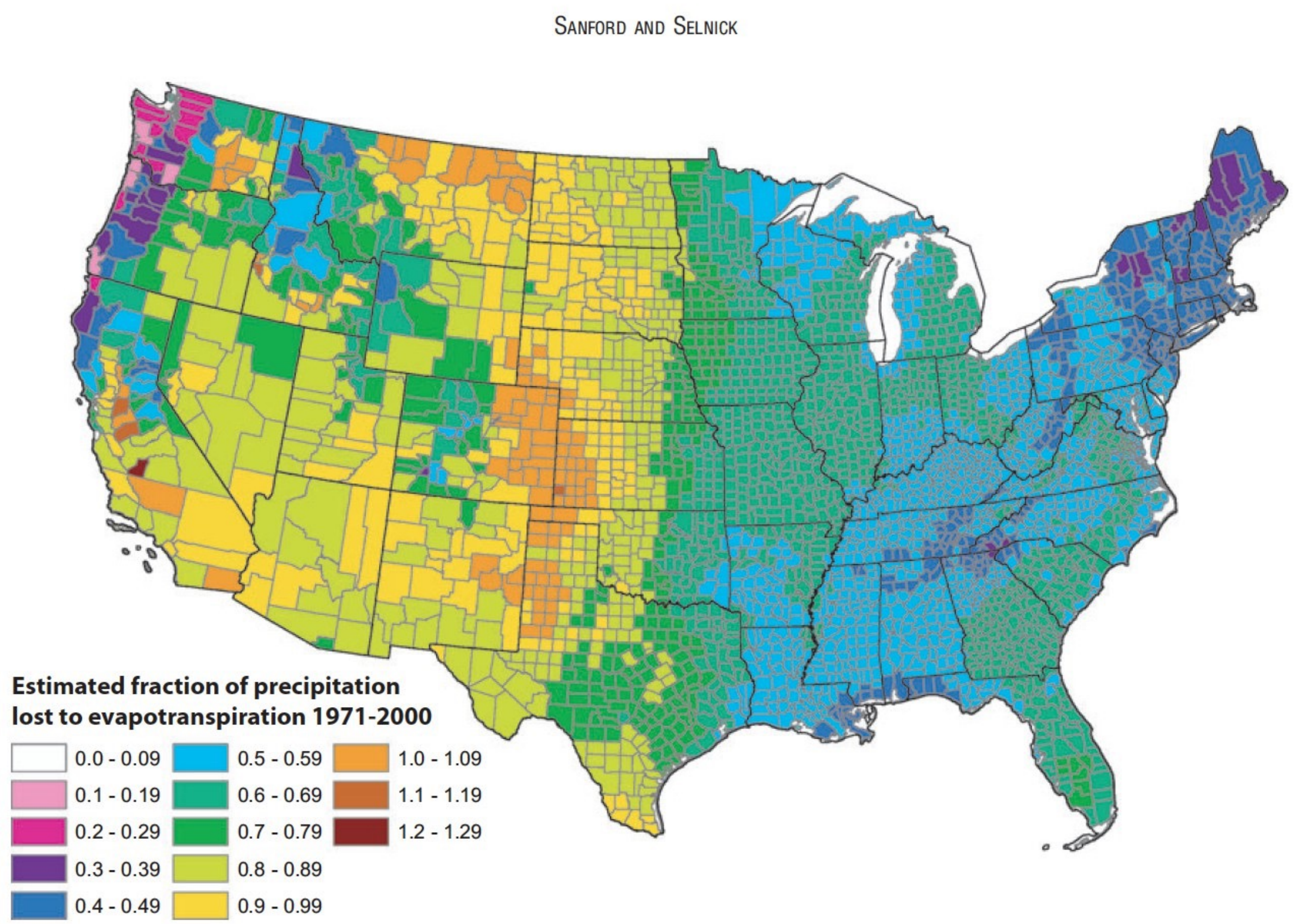
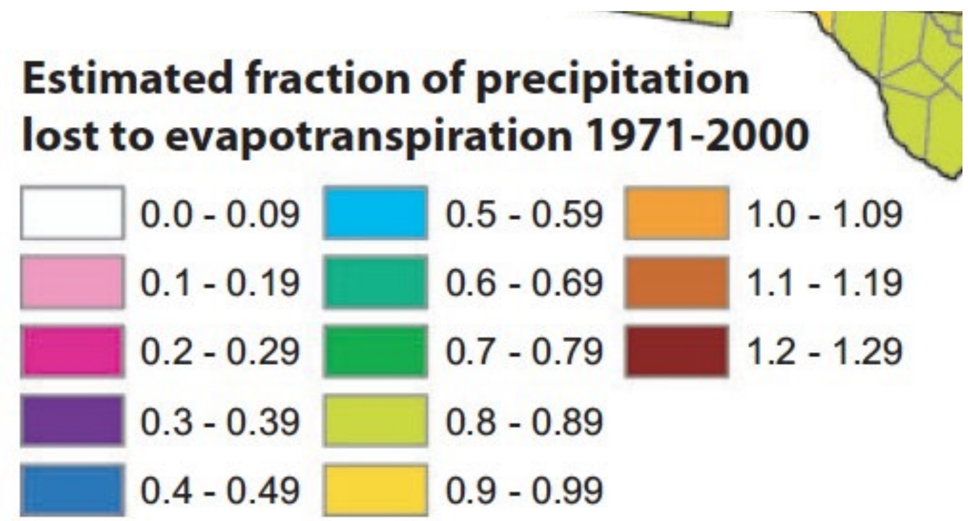


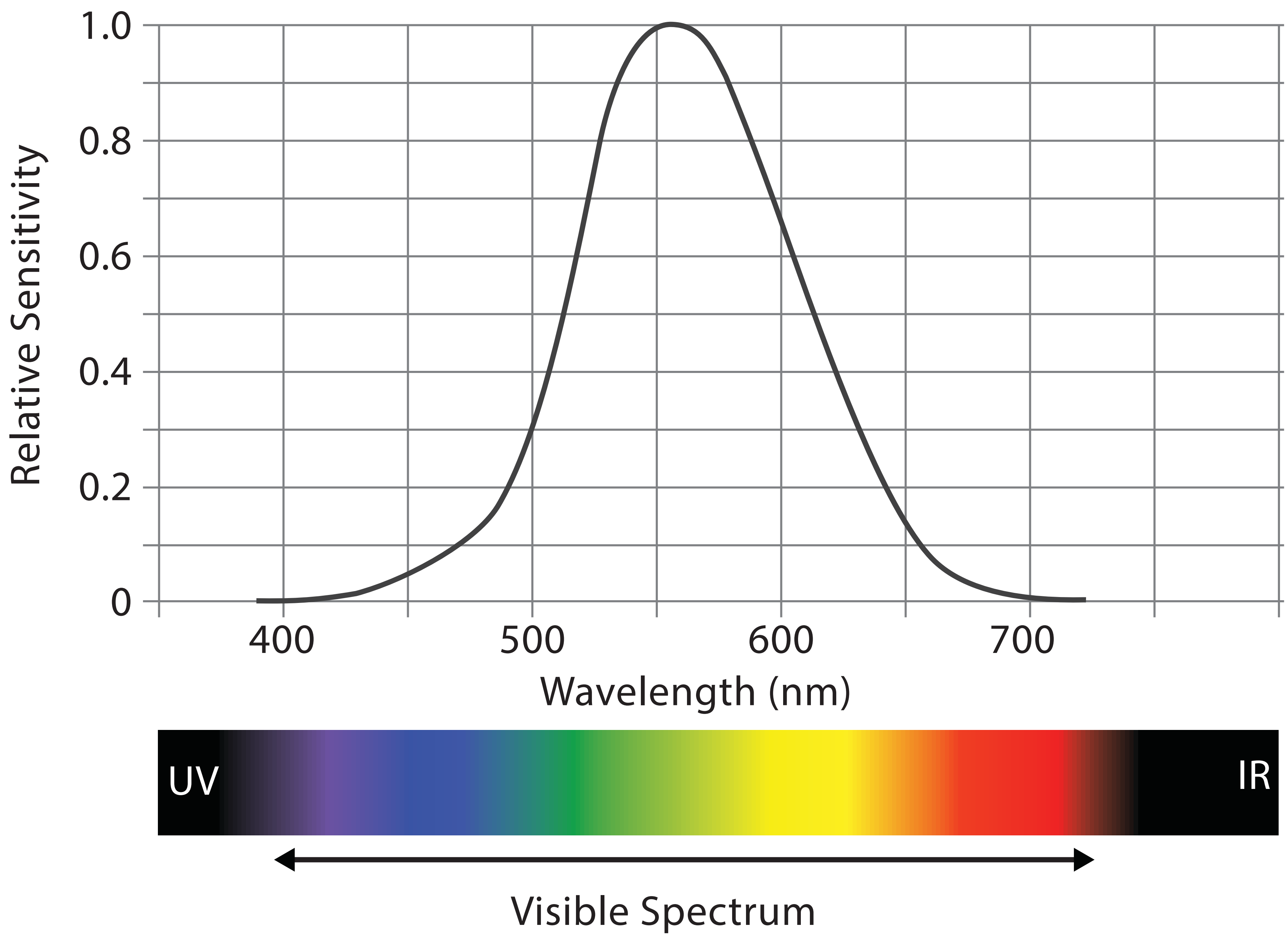
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (*P*) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/*P* were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.



The problem is that a smooth step in a value does not equate to a smooth colour transition…

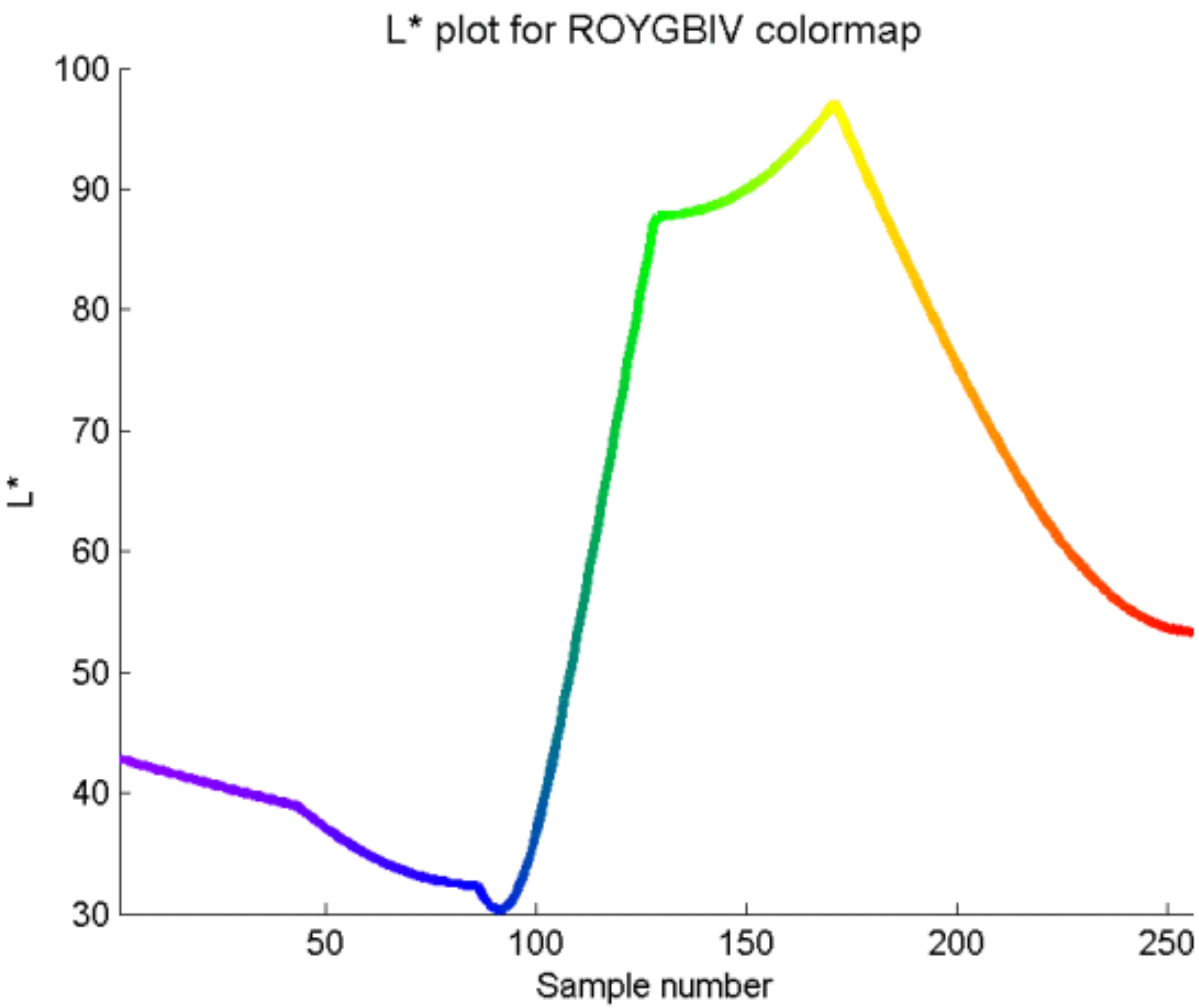
Colour

Additionally, colour is not equally binned in reality. We perceive colours differently due to an increased sensitivity to the yellow part of the spectrum...



Colour

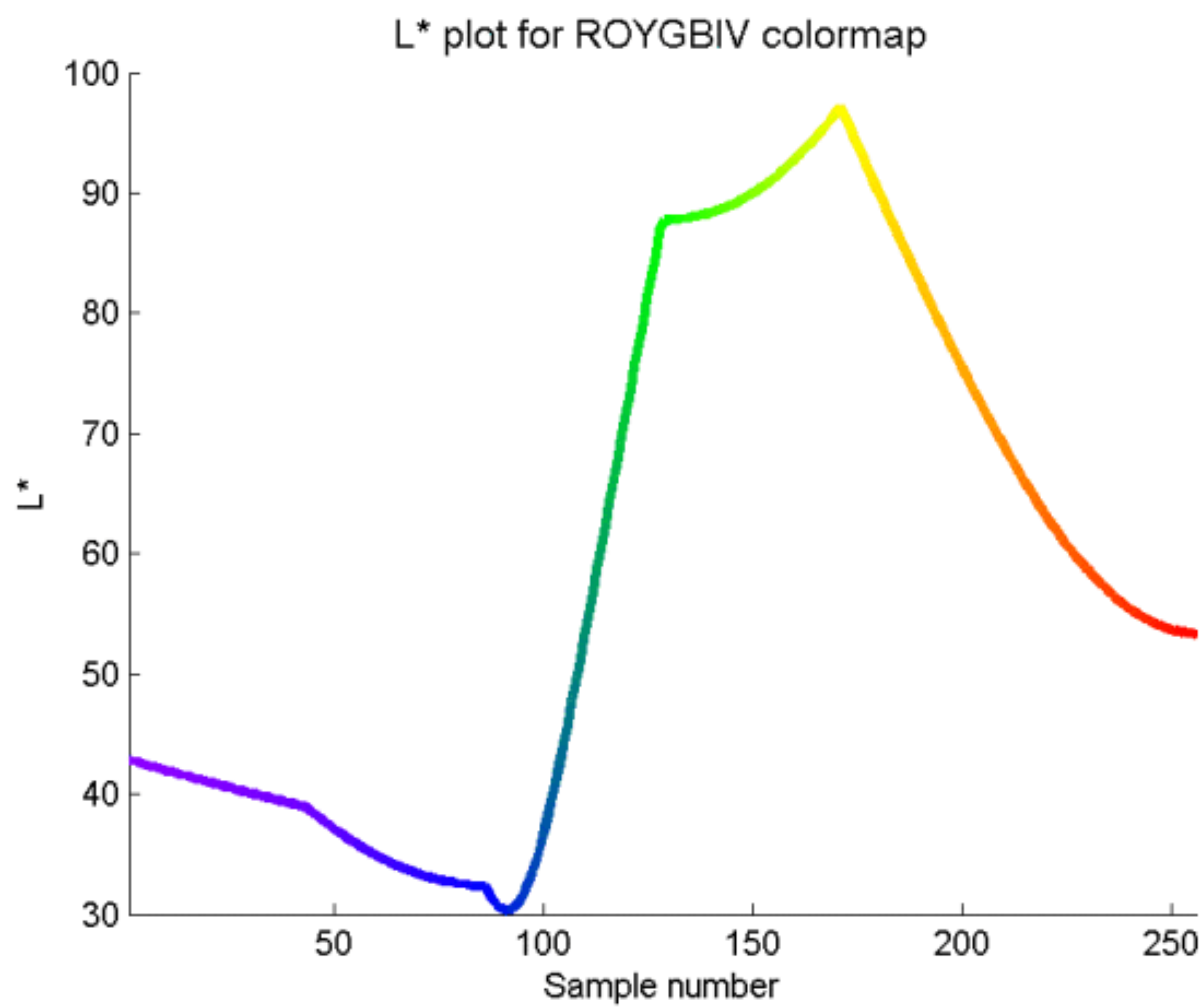
Luminosity is also not stable across the colours, meaning some colours will pop out more than others… and not always intentionally.

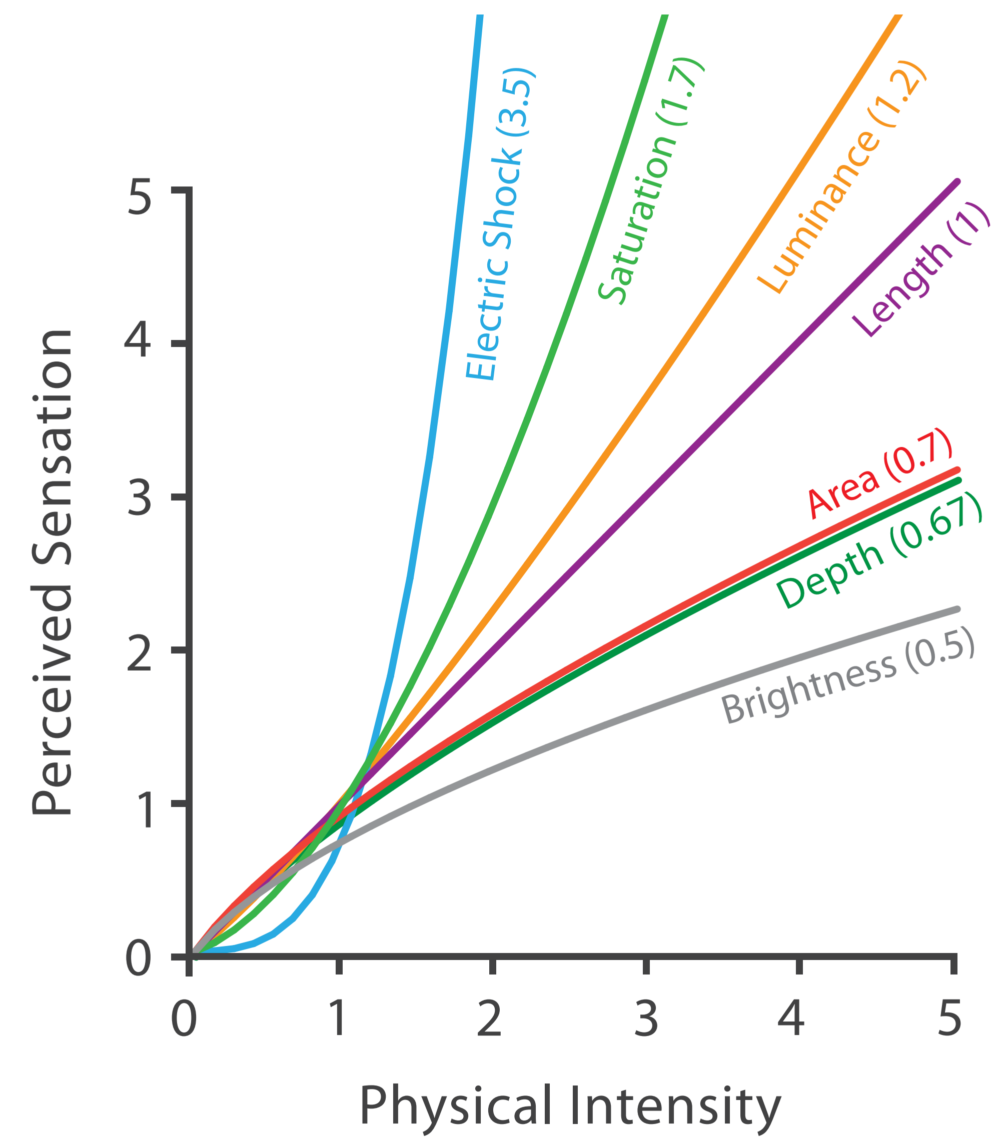


https://mycarta.wordpress.com/2012/10/06/the-rainbow-is-deadlong-live-the-rainbow-part-3/

Colour

Luminosity is also not stable across the colours, meaning some colours will pop out more than others… and not always intentionally.

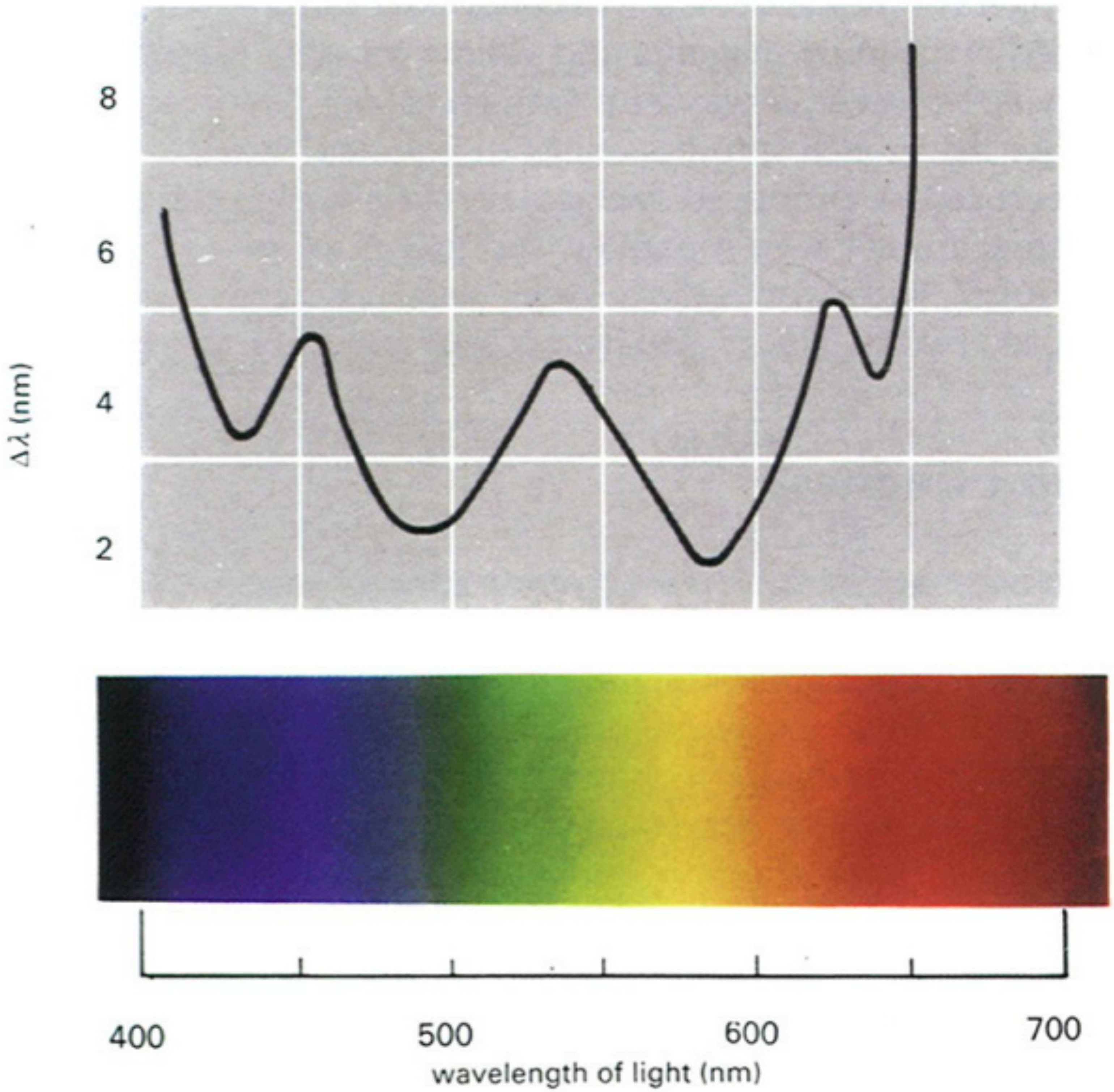




https://mycarta.wordpress.com/2012/10/06/the-rainbow-is-deadlong-live-the-rainbow-part-3/

Colour

And how we perceive changes in hue is also very different.



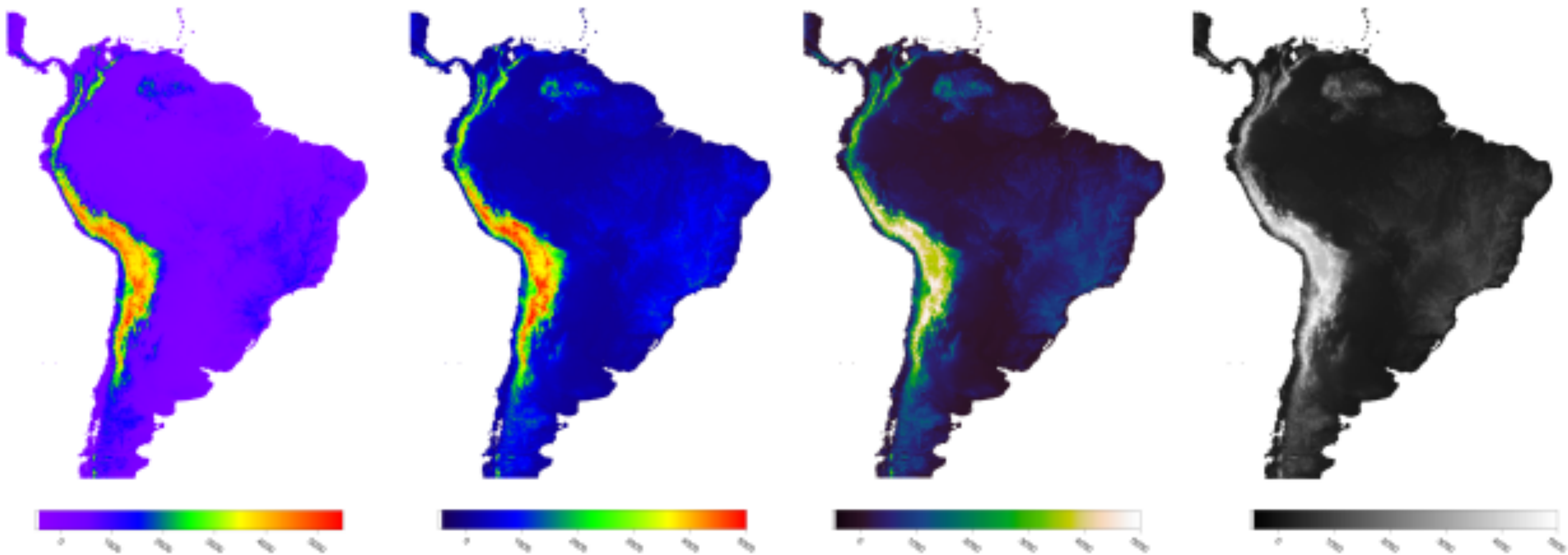
Gregory compared the wavelength of light with the smallest observable difference in hue (expressed as wavelength difference).

As you can see, the line is not flat.

Is there a colour palette for scientific visualisation that works?

Colour

HSL linear L rainbow palette



https://mycarta.wordpress.com/2012/10/06/the-rainbow-is-deadlong-live-the-rainbow-part-3/

Kindlmann, G. Reinhard, E. and Creem, S., 2002, Face-based Luminance Matching for Perceptual Colormap Generation, IEEE Proceedings of the conference on Visualization ’02

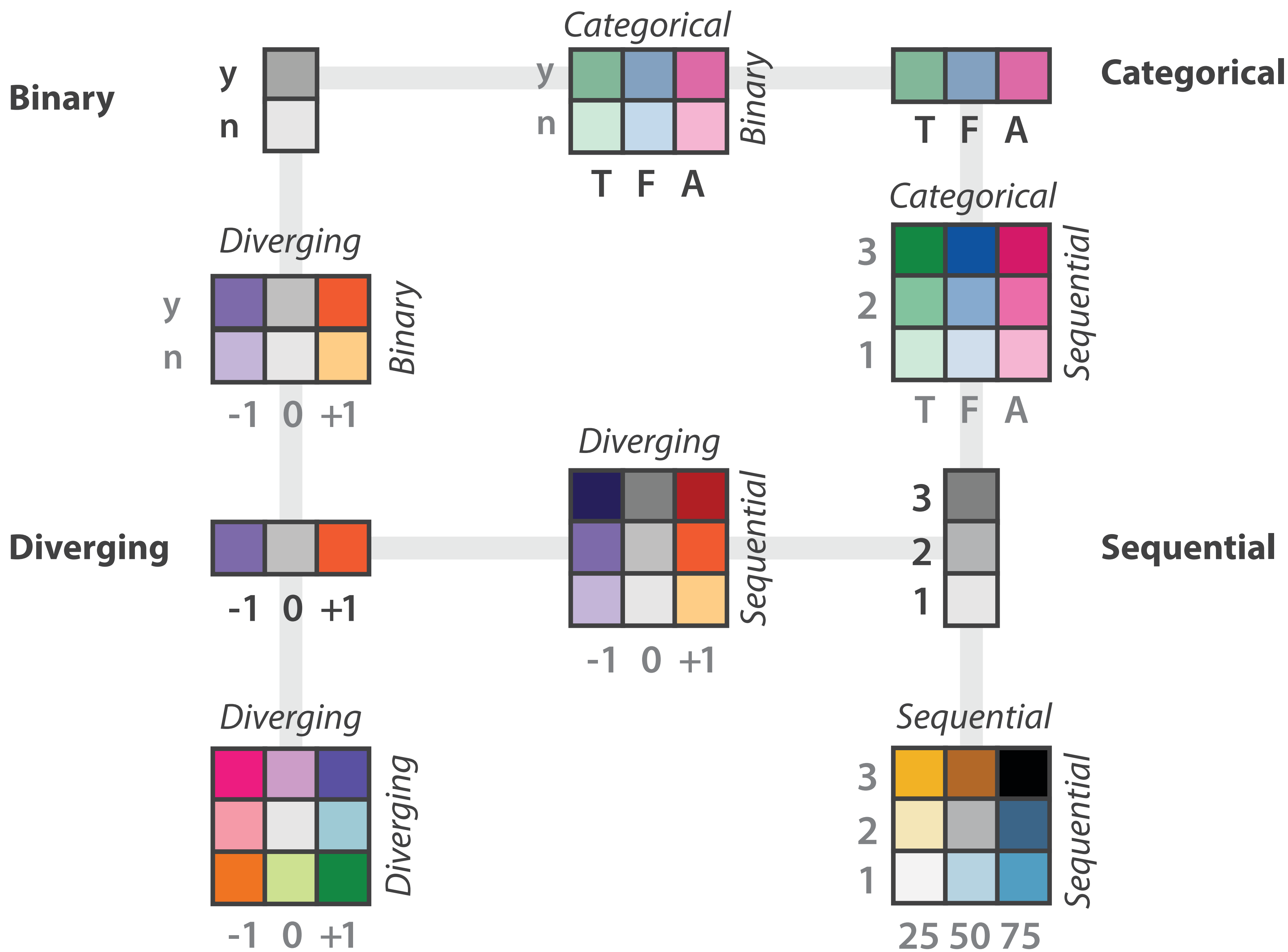
Colour

HSL linear L rainbow palette

These are available in matplotlib and therefore in seaborn, etc, so there's no excuse :)

Colour

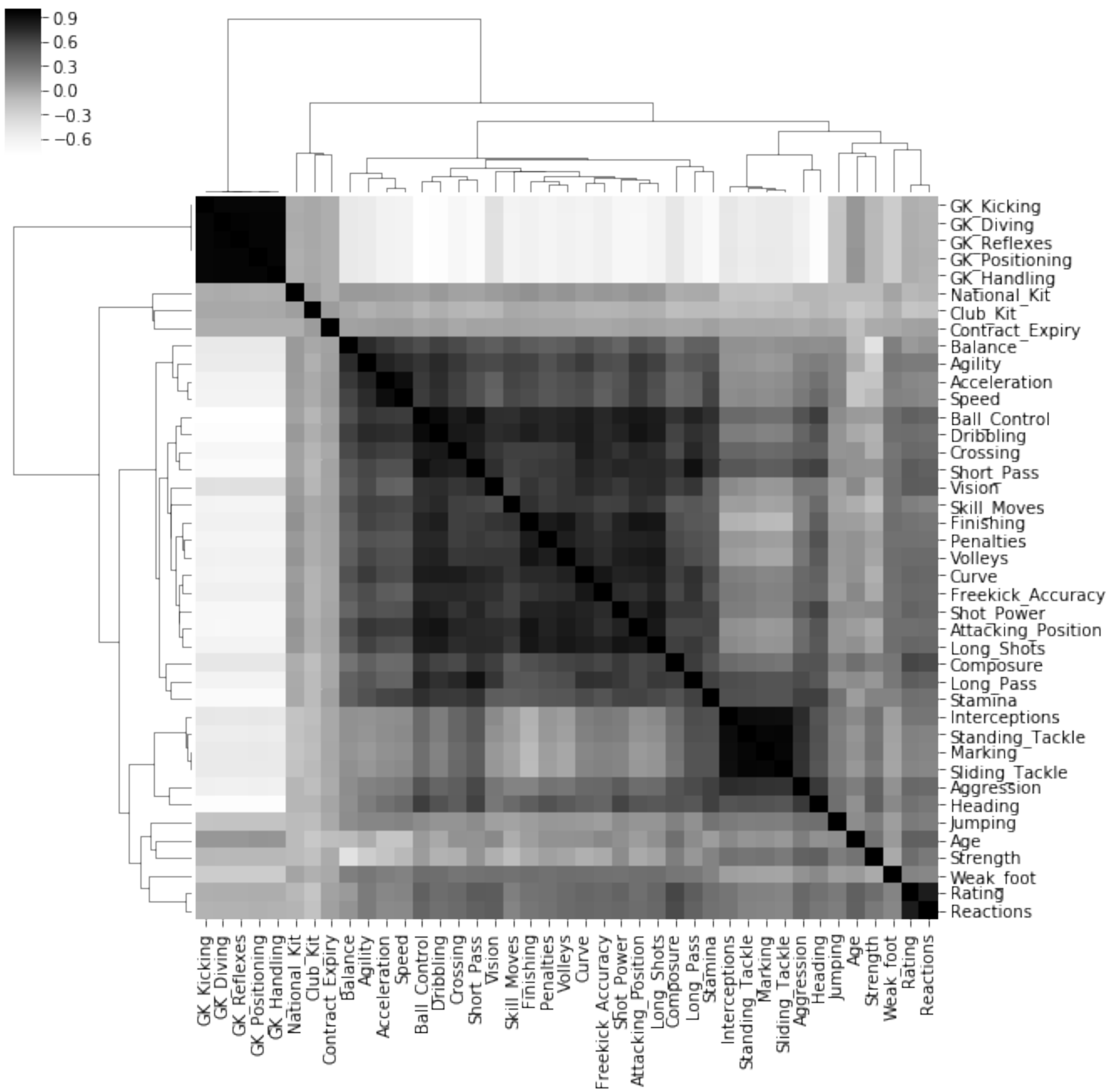
There are also lots of default colour maps that can be applied to particular data types.

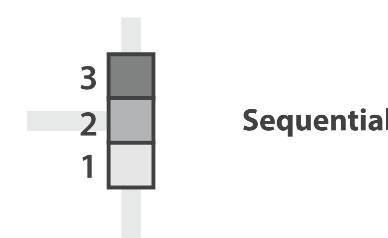


http://colorbrewer2.org/

Color

Here I'm showing the correlation between football player attributes.
Is the choice of colour map helping this comparison?

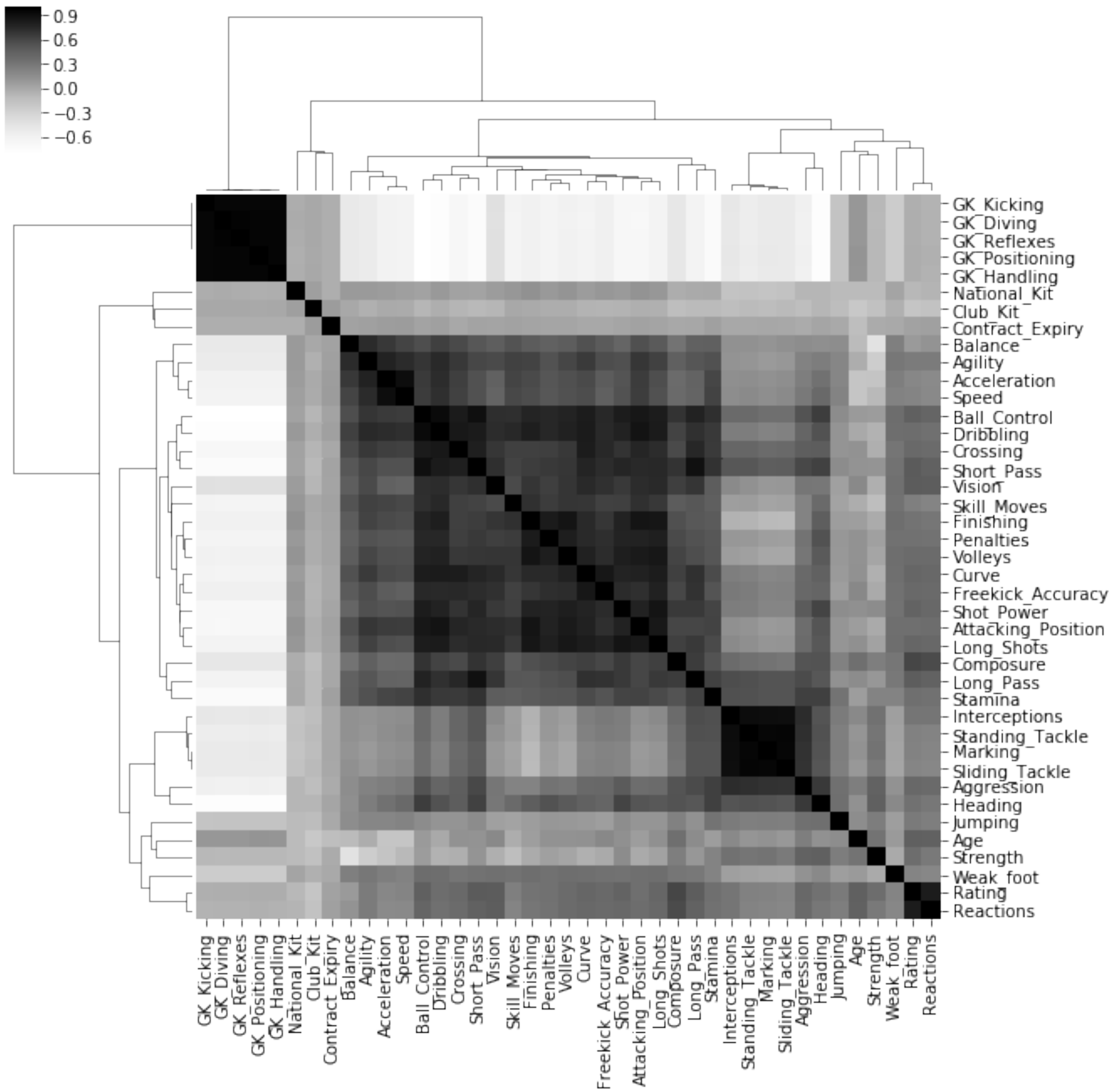


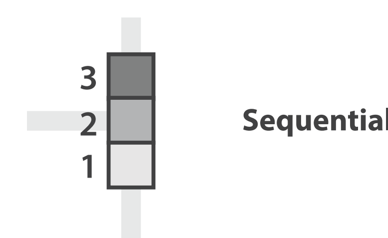



```
import seaborne as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

Color

Here I'm showing the correlation between football player attributes.
Is the choice of colour map helping this comparison?



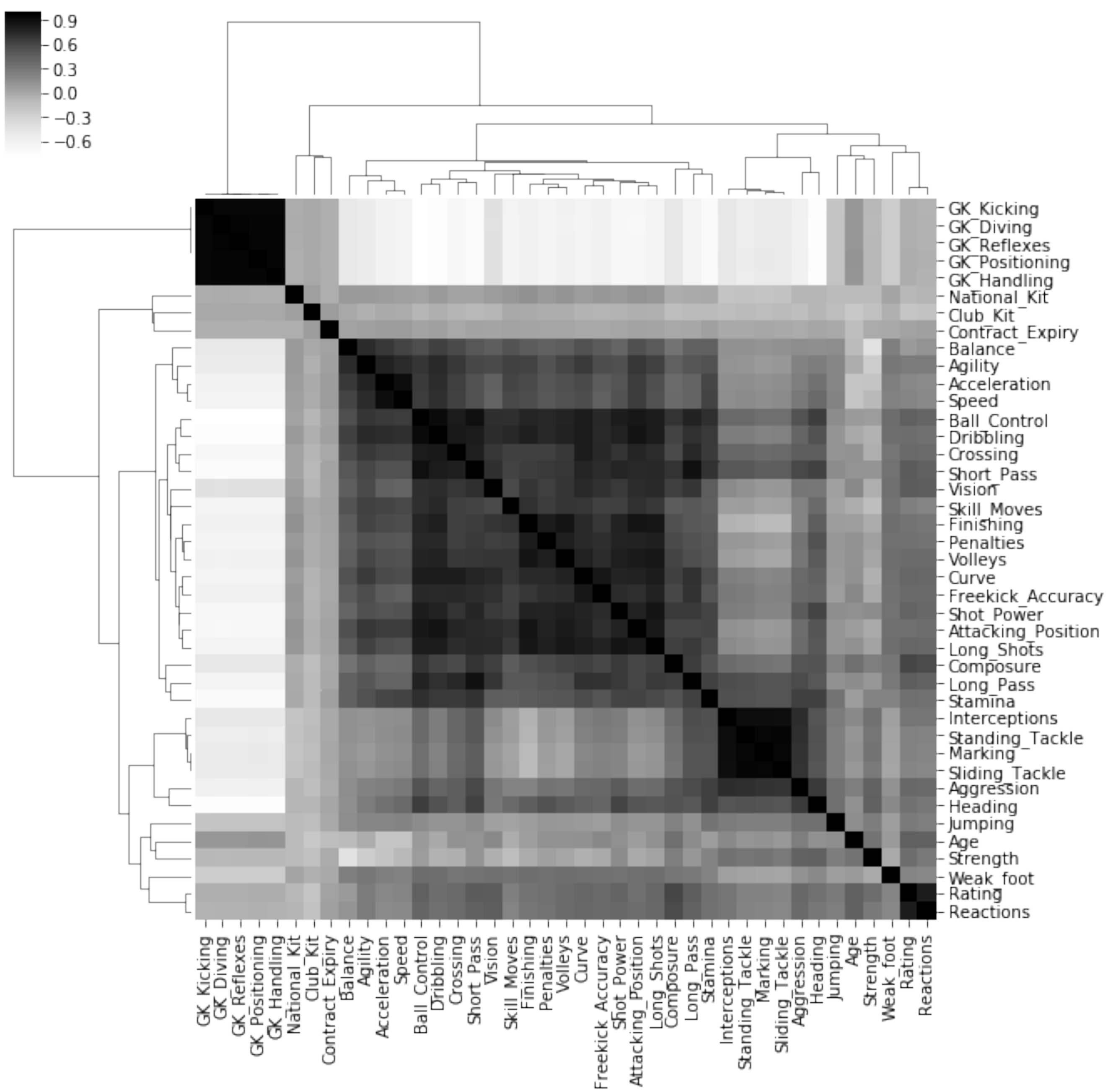


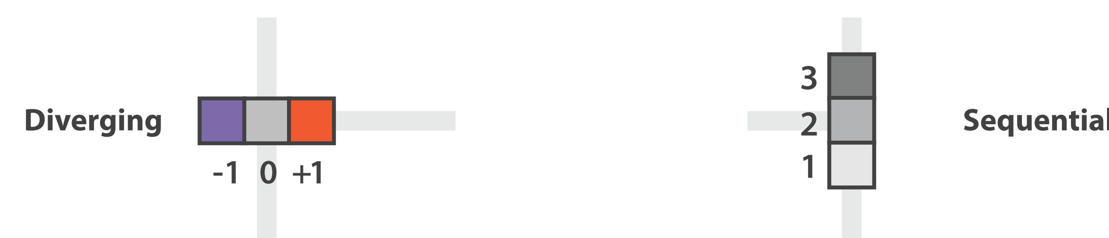
But our value pivots around 0, so the scale should be a diverging one.

```
import seaborne as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

Color

Here I'm showing the correlation between football player attributes. Is the choice of colour map helping this comparison?



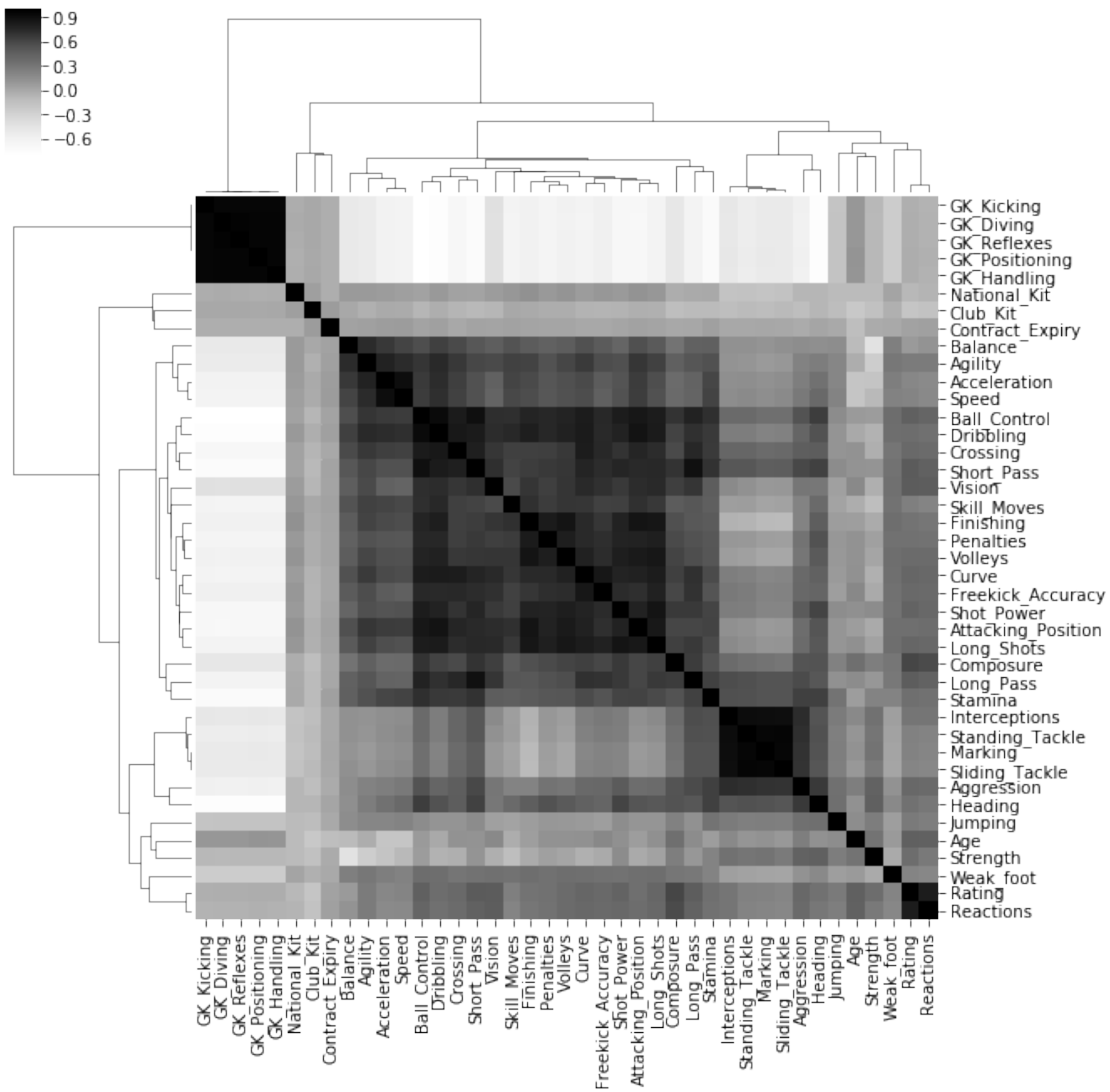


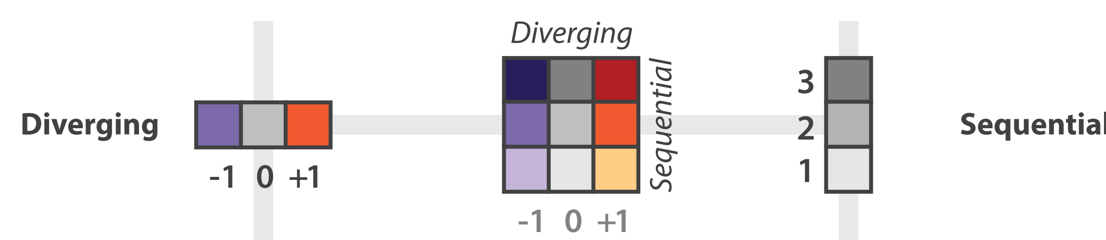
But our value pivots around 0, so the scale should be a diverging one.

```
import seaborne as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

Color

Here I'm showing the correlation between football player attributes. Is the choice of colour map helping this comparison?



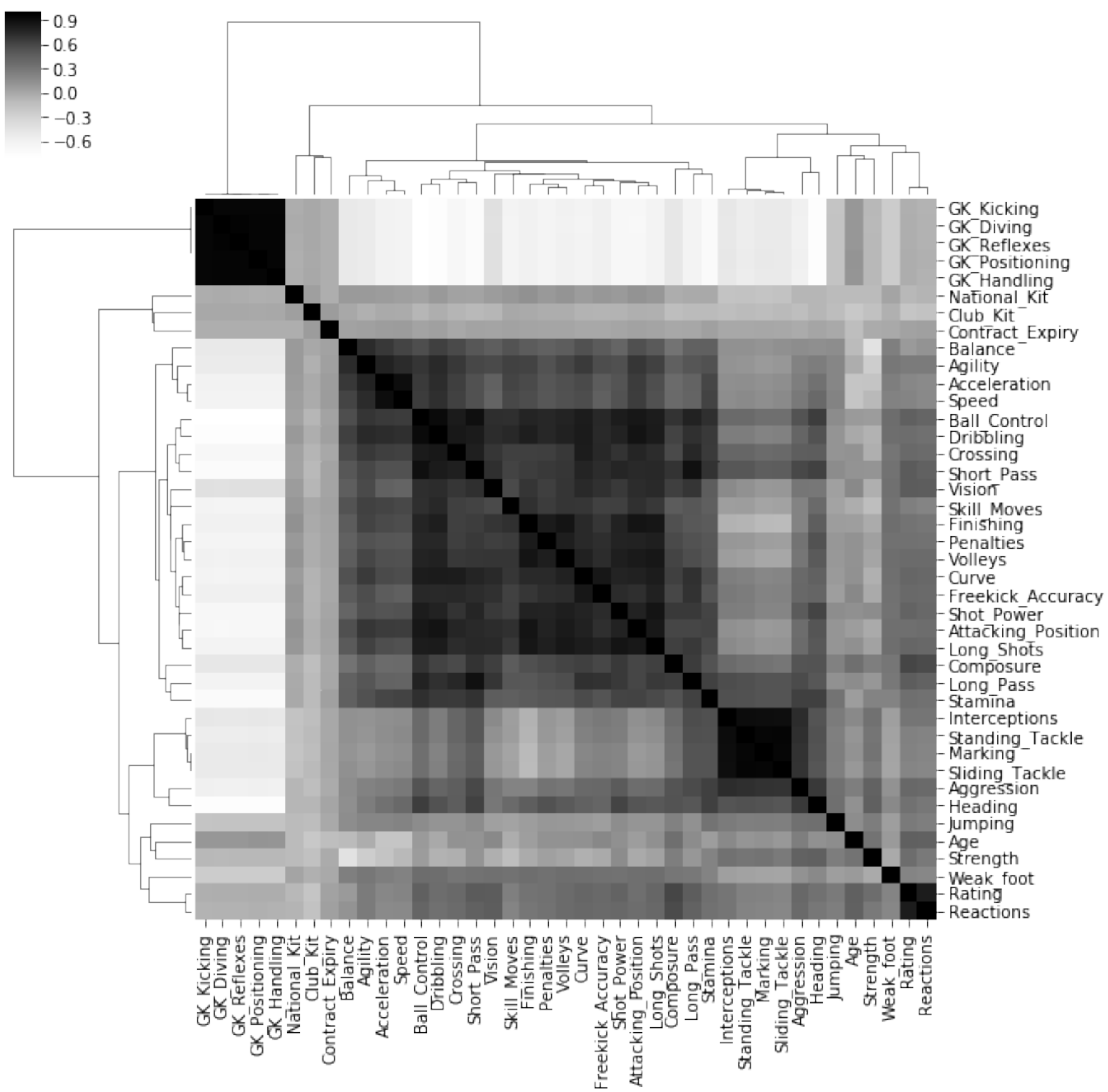


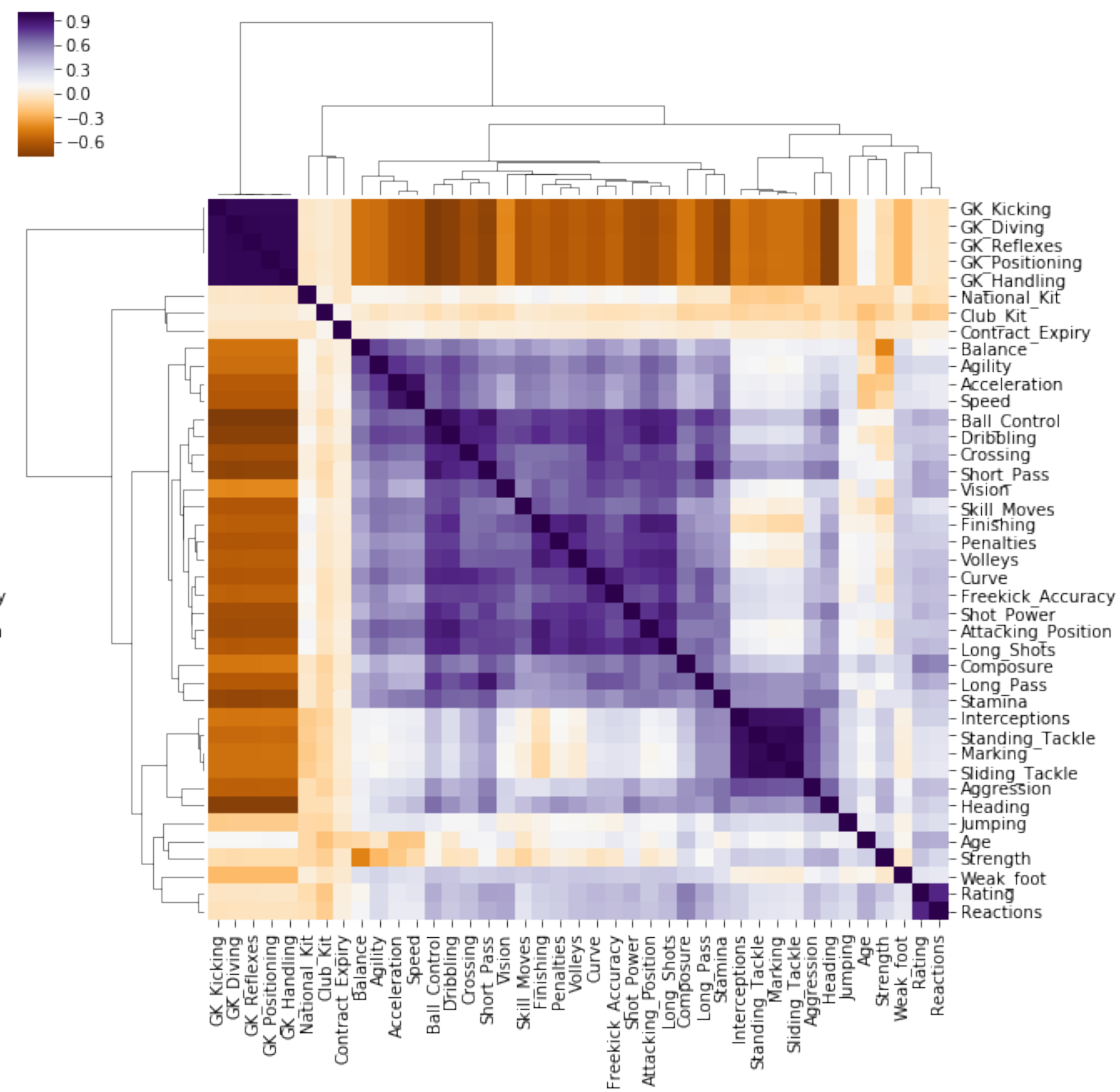
But our value pivots around 0, so the scale should be a diverging one.

```
import seaborne as sns
sns.clustermap(fifa.corr(), cmap='Greys')
```

Color

Here I'm showing the correlation between football player attributes. Is the choice of colour map helping this comparison?





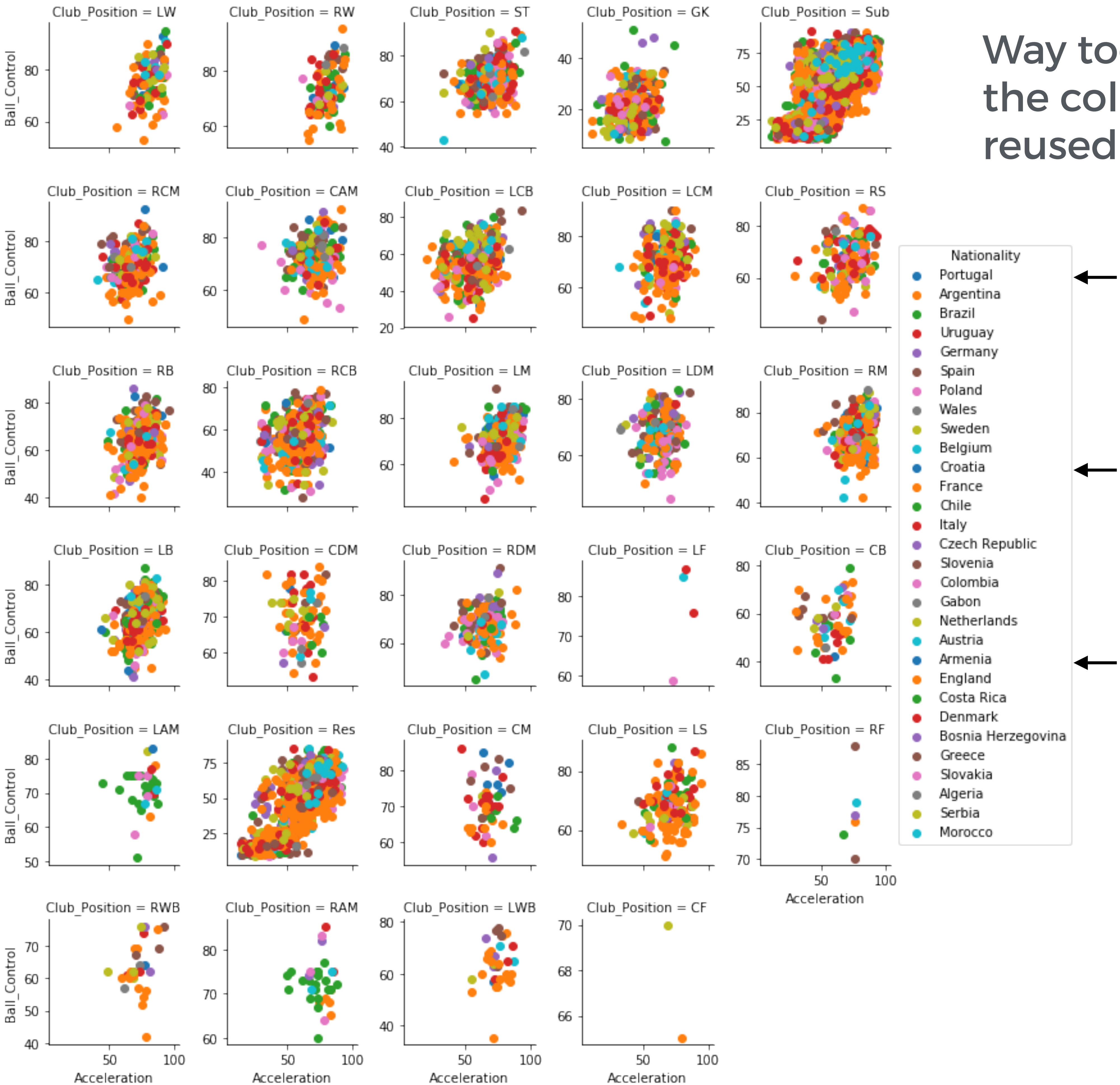

```
import seaborn  as sns
sns.clustermap(fifa.corr(), cmap='PuOr')
```

Color

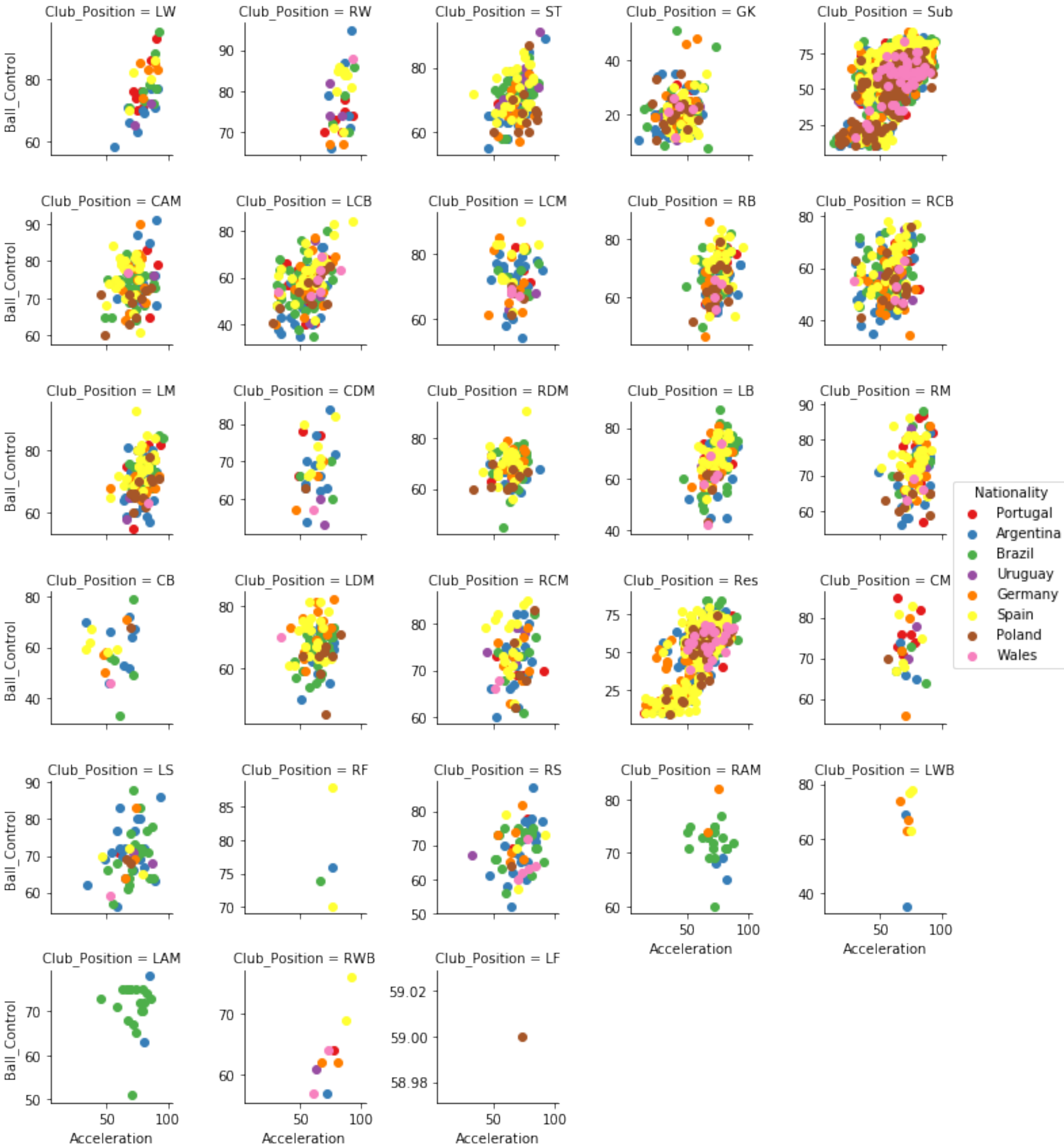
You also don’t want to have too many colours.

Too many colours means that users have to remember what a colour means. So a max of around 8 categories in a plot is recommended, otherwise the ‘distance’ between colours becomes too small.

Way too many...and the colours get reused.



Here 10 colours are being used to represent 29 countries.



Much better

Here 8 colours are being used to represent 8 countries.