Eamonn Maguire

Principles of Data Visualization I
University of Liege, Belgium
27th April 2018

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

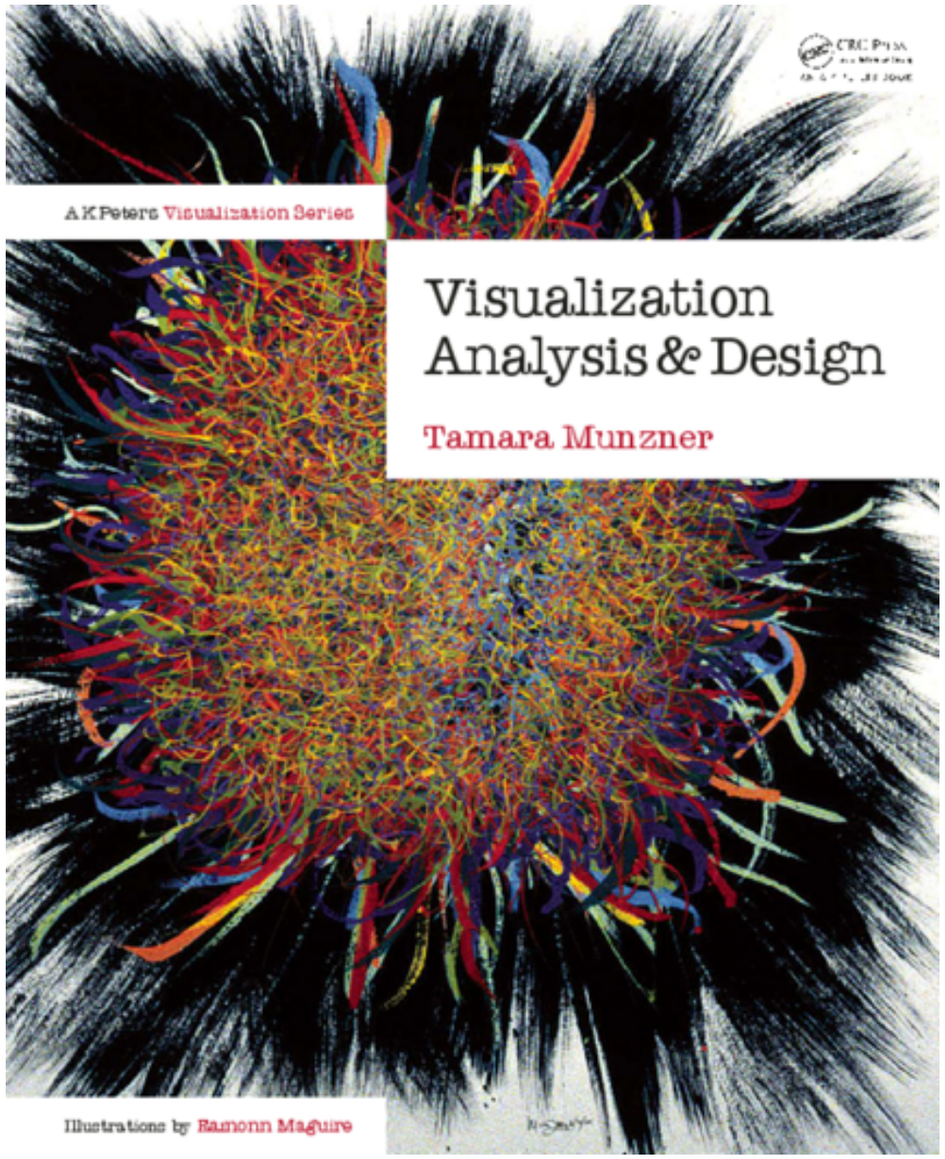
* from Noel Illinsky, http://complexdiagrams.com/

Lecture outcomes

The **what**? Major data types and classifications of them

The **why**? Why are we visualising at all?

The **how**? How can we visualize? What archetypes can we use to guide us?



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.

Data Panel

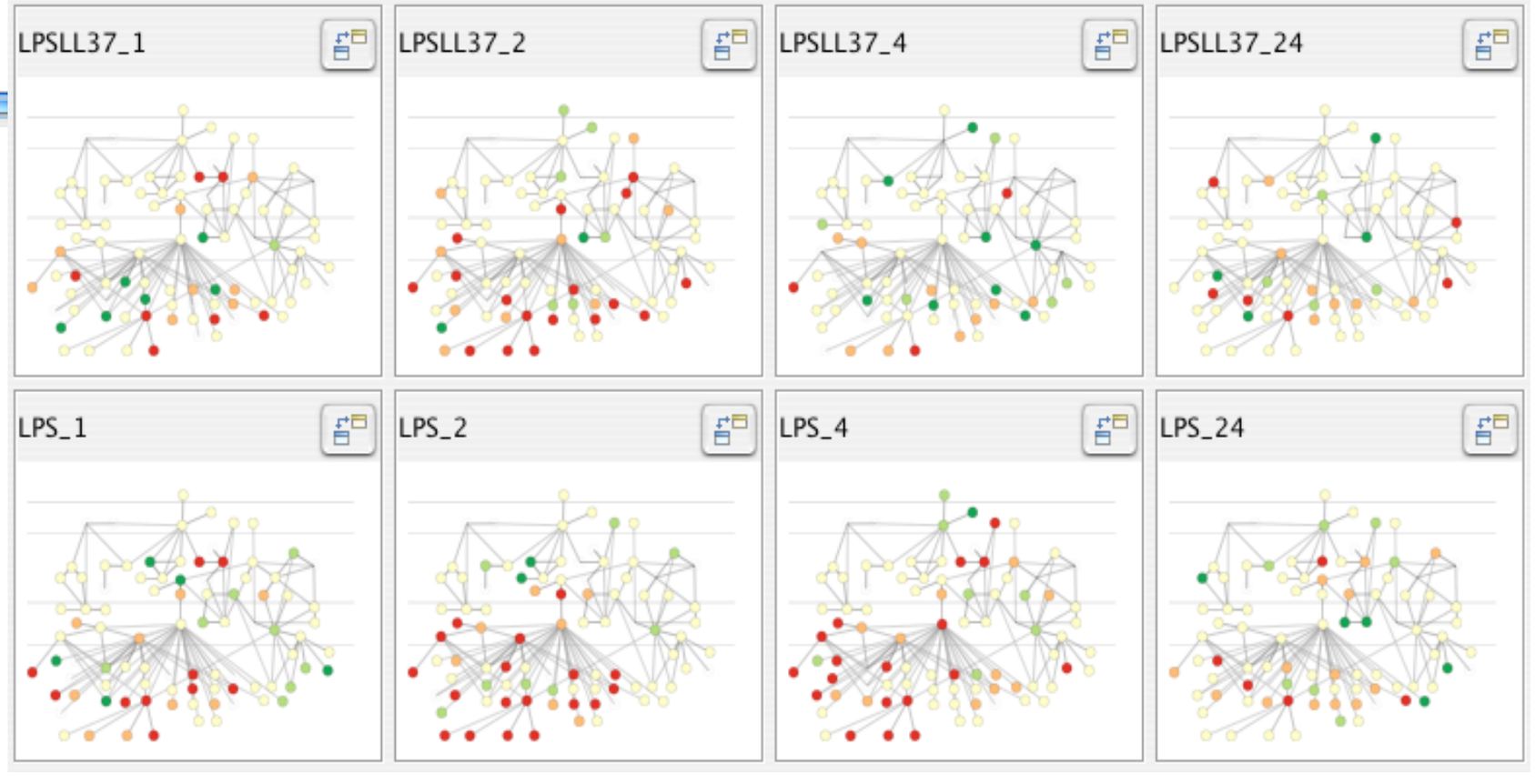
ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_24	LPSLL37_24_pvals
IRAK2	Kinase	2.367	0.251	1.337	-1.553	
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303	0.807
CXCL2	Chemokine	1.853	0.376	4.111	-1.019	0.745
CHUK	Kinase	-1.376	0.373	2.232	1.194	0.387
IL13	Cytokine	-5.961		2.139	-1.236	0.601
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			
CCL11	Chemokine	-1.338	0.349			
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801			

Expression color scale

-2.5 0 2.5

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.]



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	

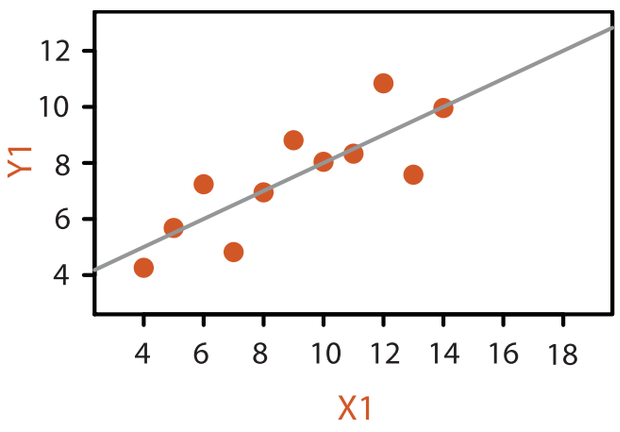
Why visualize?

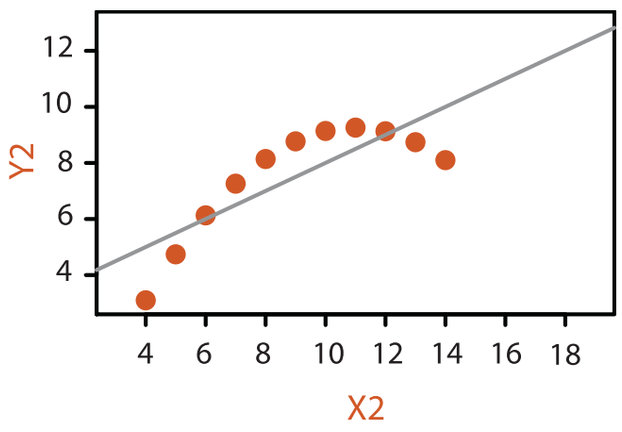
The role of visualization systems is to provide visual representations of datasets that help people carry out tasks more effectively.

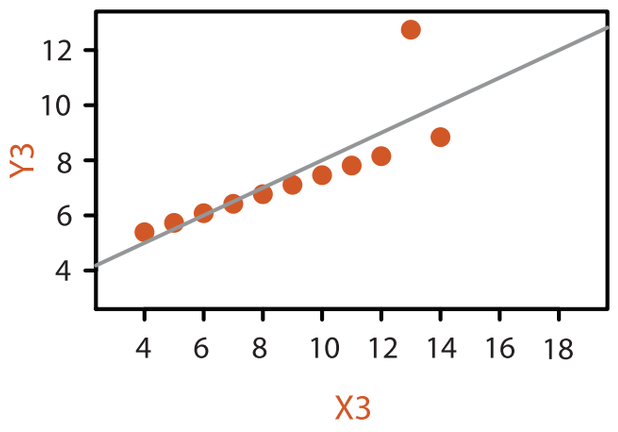
Anscombe's Quartet: Raw Data

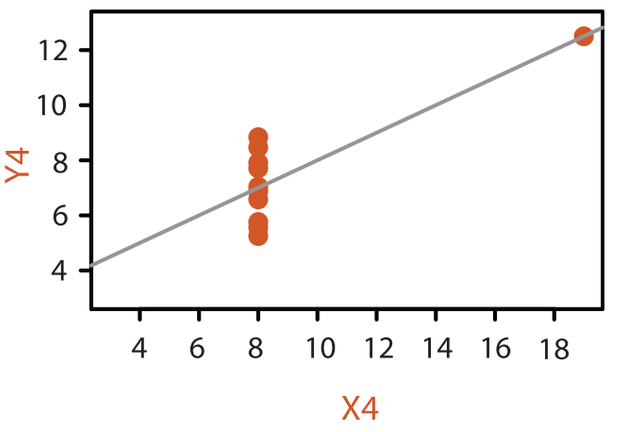
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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



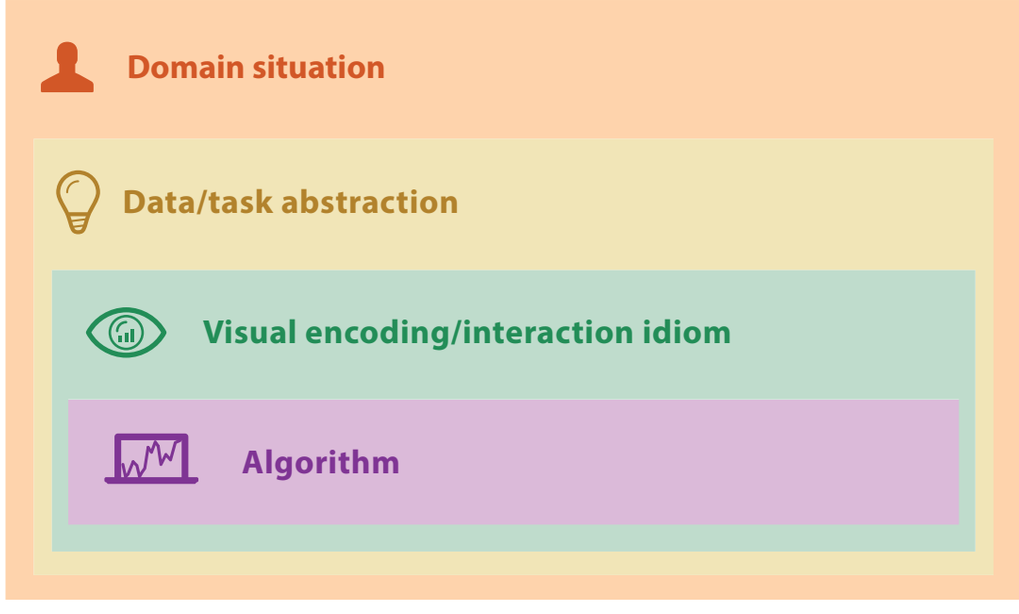






Analysis framework: Four levels, three questions

- Domain situation
  - who are the target users?
- Data/Task Abstraction
  - translate from specifics of domain to vocabulary of vis
    - What is shown? **Data abstraction**
    - Why is the user looking at it? **Task abstraction**
- Visual Encoding
  - How is it shown?
    - **visual encoding**: how to draw
    - **interaction**: how to manipulate
- Algorithm
  - efficient computation, layout algorithms etc.



A Multi-Level Typology of Abstract Visualization Tasks

Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).

A Nested Model of Visualization Design and Validation.

Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).

What are you visualising?

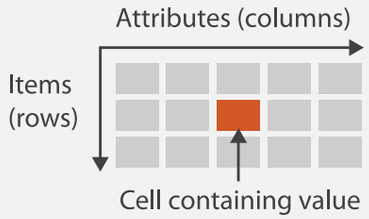
DATA TYPES

STATIC

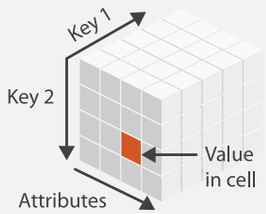
DYNAMIC

DATASET TYPES

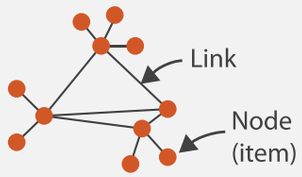
TABLES



- Multidimensional Table

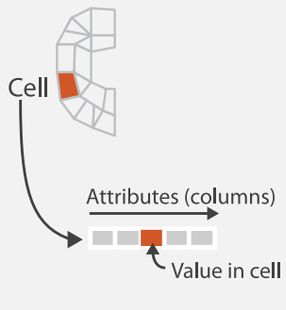


NETWORKS



- Trees

FIELDS (CONTINUOUS)



GEOMETRY (SPATIAL)

TEXT

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

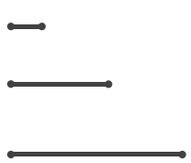
ATTRIBUTE TYPES

CATEGORICAL

ORDERED

- Ordinal

- Quantitative

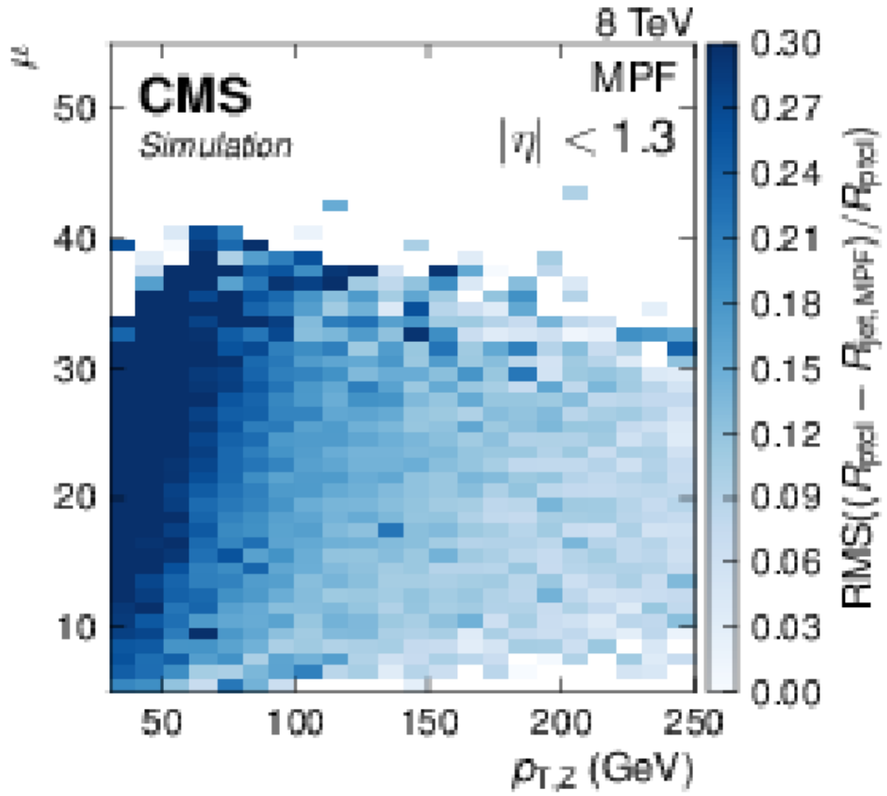


- Sequential

- Diverging

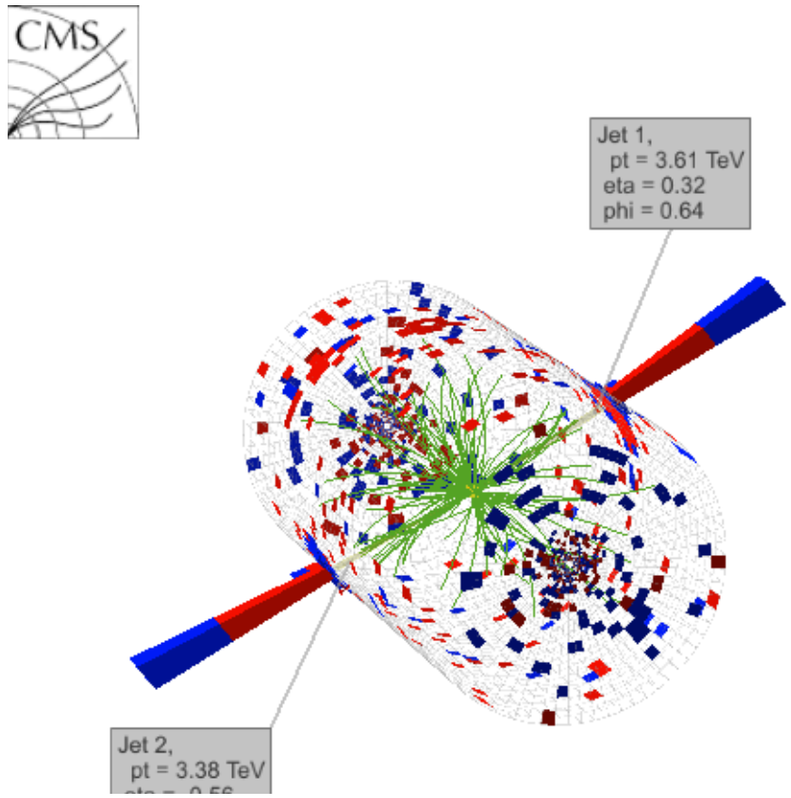
- Cyclic

The branches of data visualization



Information Visualization

Position is derived.
Incl. GeoVis



Scientific Visualization

Position is given.
Also medical visualizations

Why are you visualising the data?

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

Targets

All Data

- Trends
- Outliers
- Features

Attributes

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

Network Data

- Topology
- *Paths*

Spatial Data

- Shape

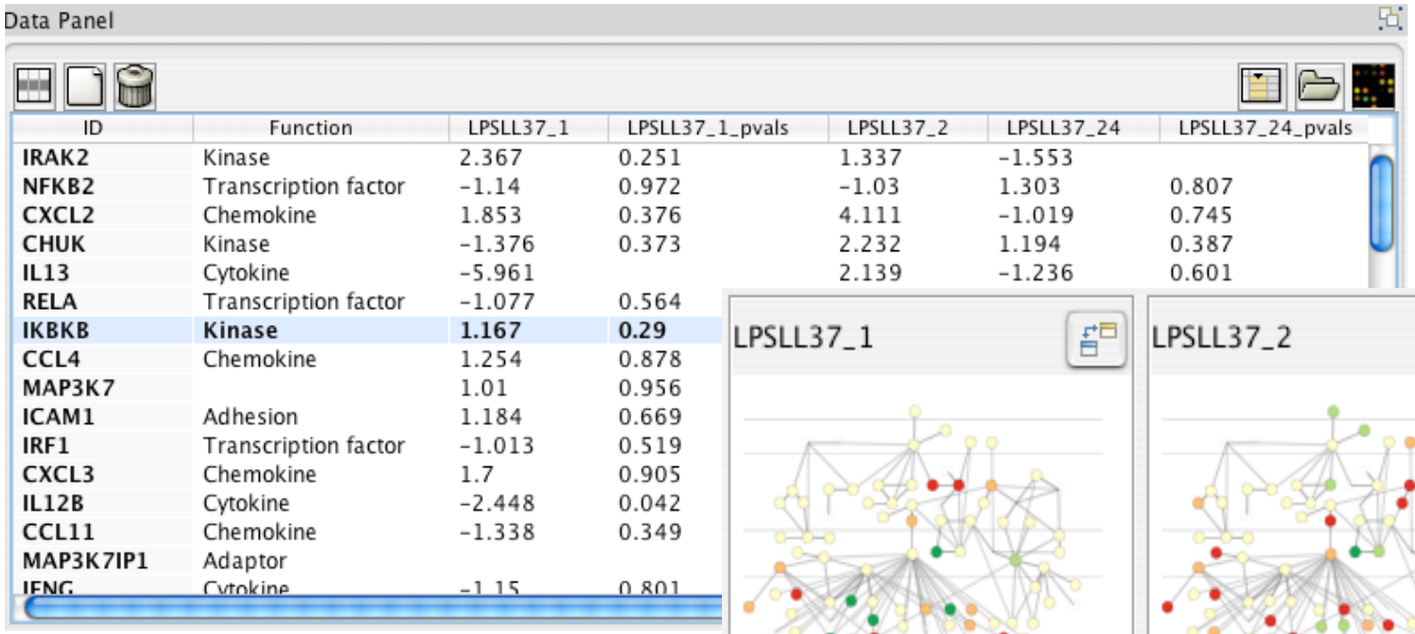
- Use
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*



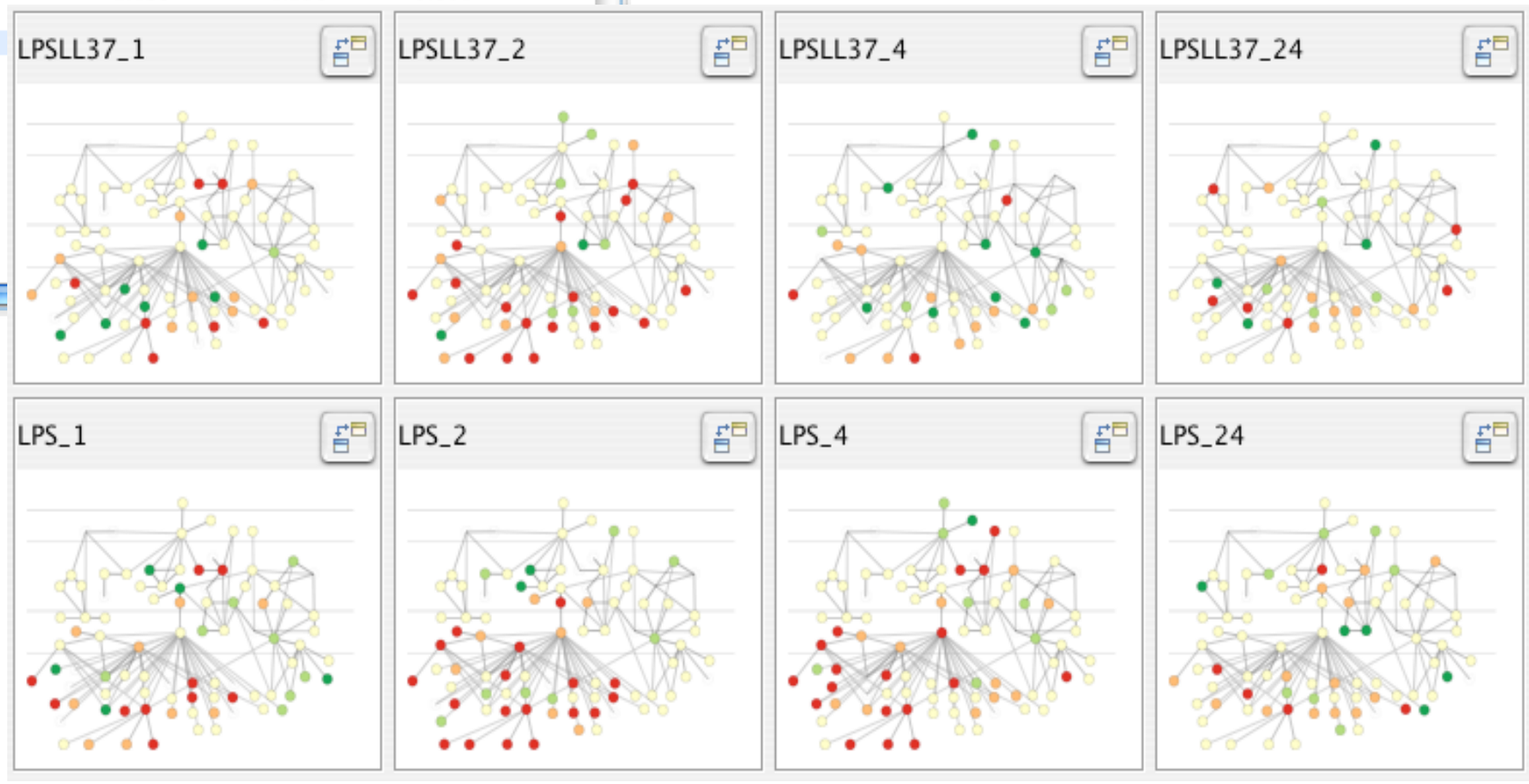
Discover

Finding new insights in your data

Implies a level of interactivity to query, compare, correlate etc.



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			
IKBKB	Kinase	1.167	0.29			
CCL4	Chemokine	1.254	0.878			
MAP3K7		1.01	0.956			
ICAM1	Adhesion	1.184	0.669			
IRF1	Transcription factor	-1.013	0.519			
CXCL3	Chemokine	1.7	0.905			
IL12B	Cytokine	-2.448	0.042			
CCL11	Chemokine	-1.338	0.349			
MAP3K7IP1	Adaptor					
IFNG	Cytokine	-1.15	0.801			



This is typically where one should be careful in how information is presented.
An erroneous data encoding could bring about wrong conclusions.

Use

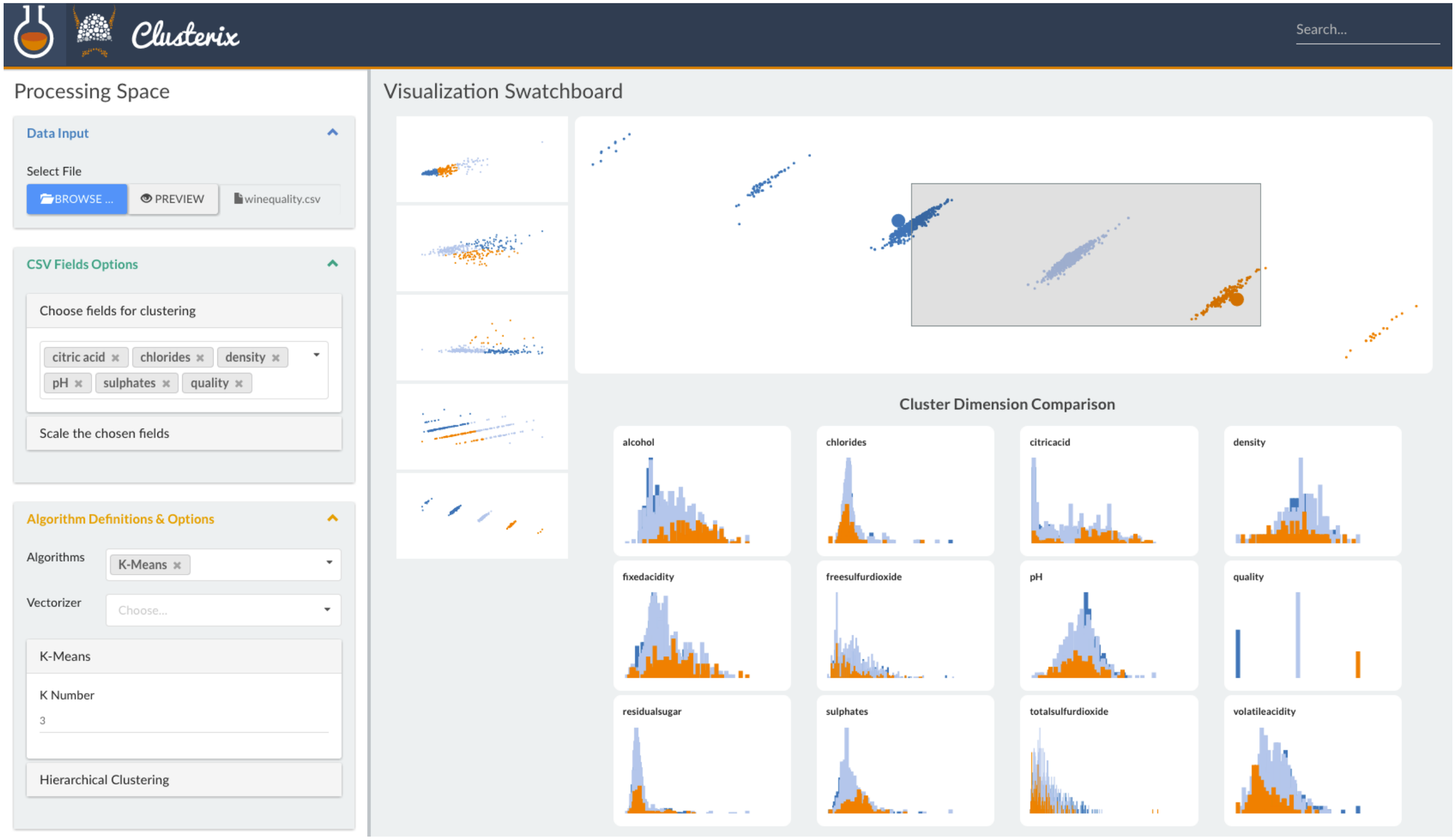
→ Consume

→ *Discover* → *Present* → *Enjoy*

Discover

Finding new insights in your data

Implies a level of interactivity to query, compare, correlate etc.



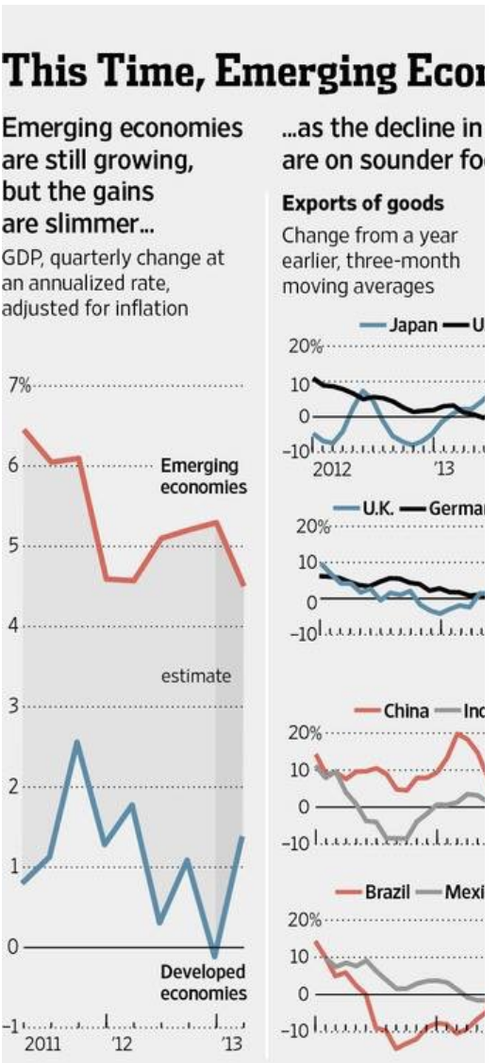
Work with Ilias Koutsakis and Gilles Louppe

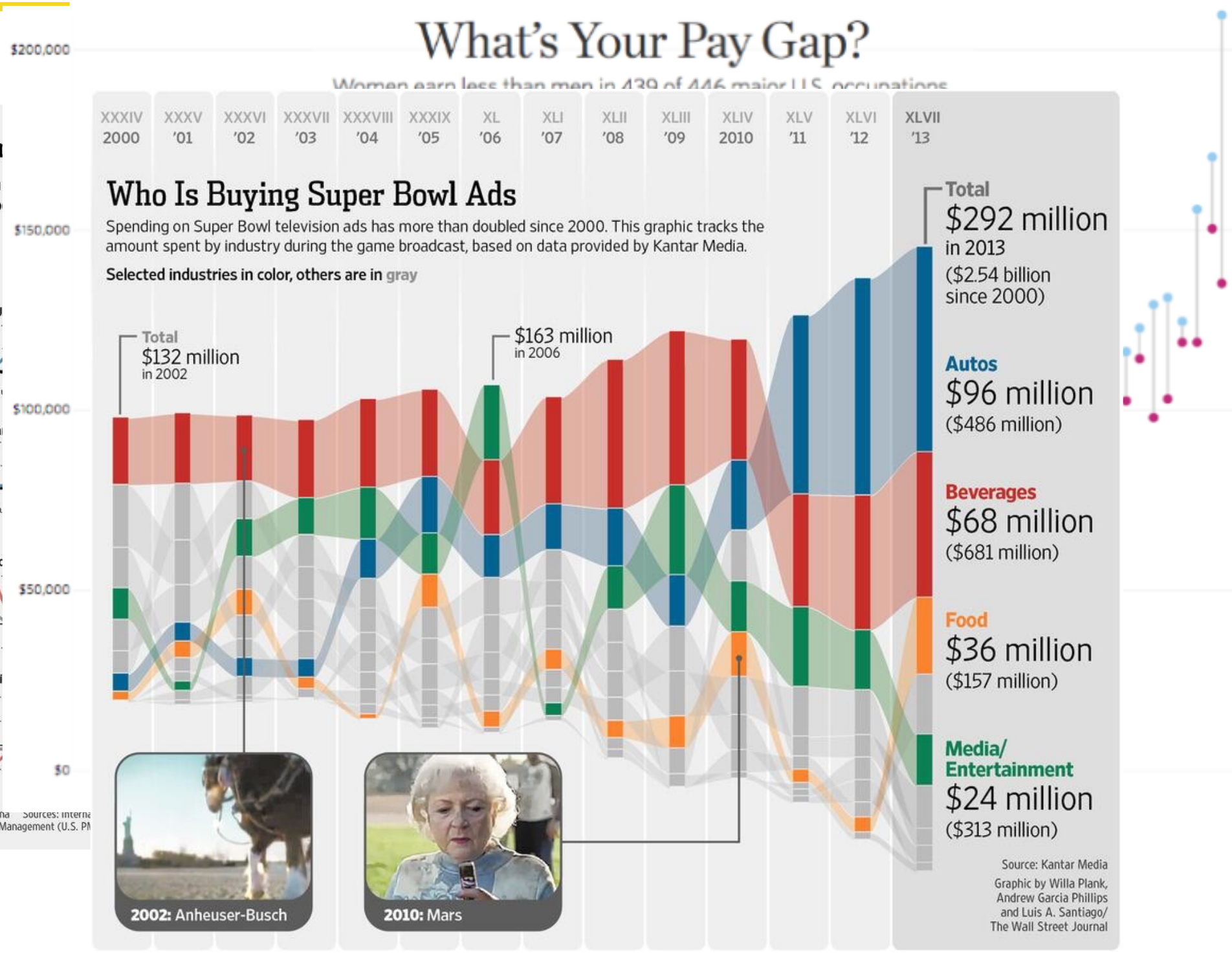
Visual Analytics

- Use
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*


Present

Presenting your results, e.g. for a paper





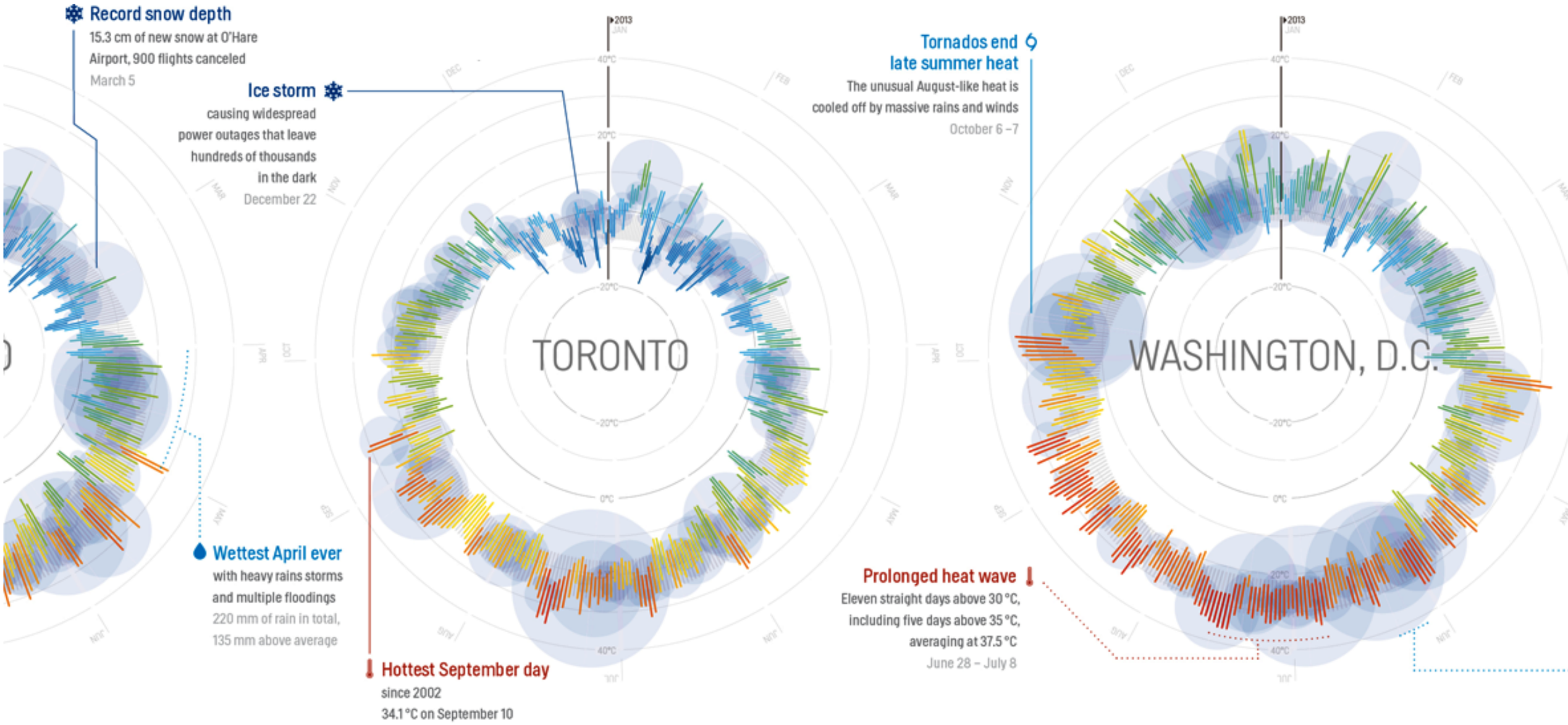
[From Wall Street Journal Graphics]

- Use
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*



Present

Presenting your results, e.g. for a paper

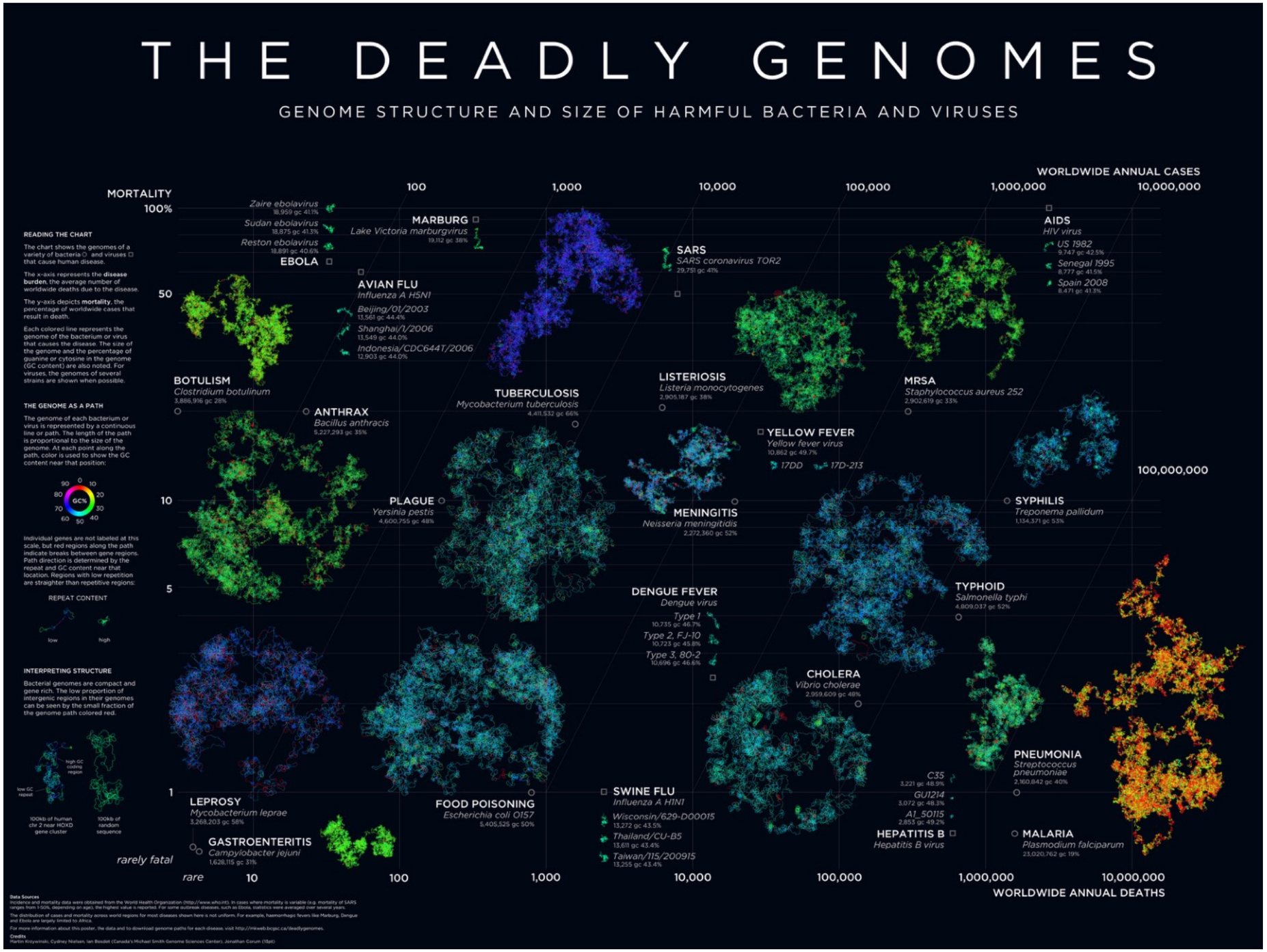


From http://weather-radials.com/

- Use
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*

Present

Presenting your results, e.g. for a paper

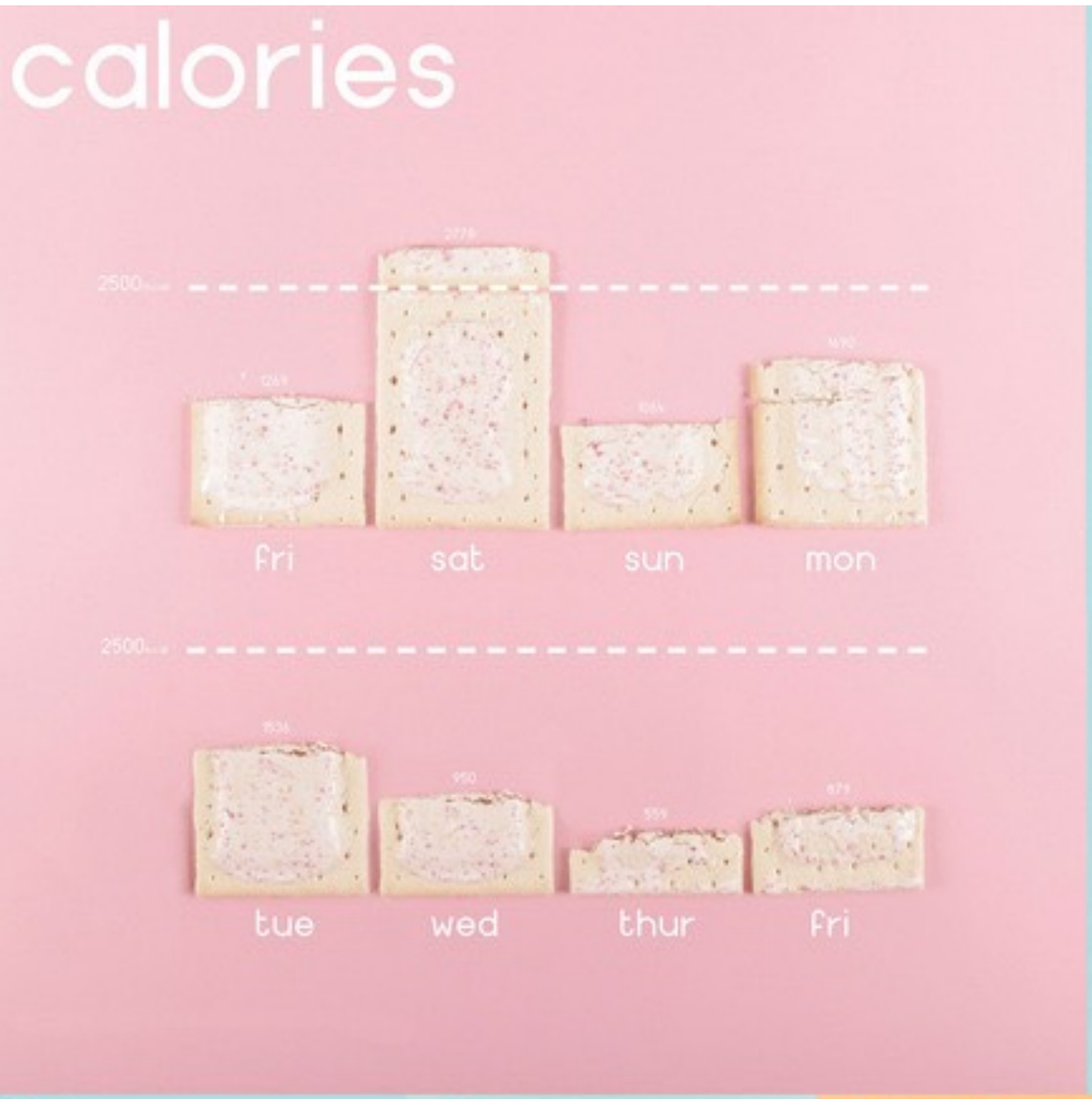


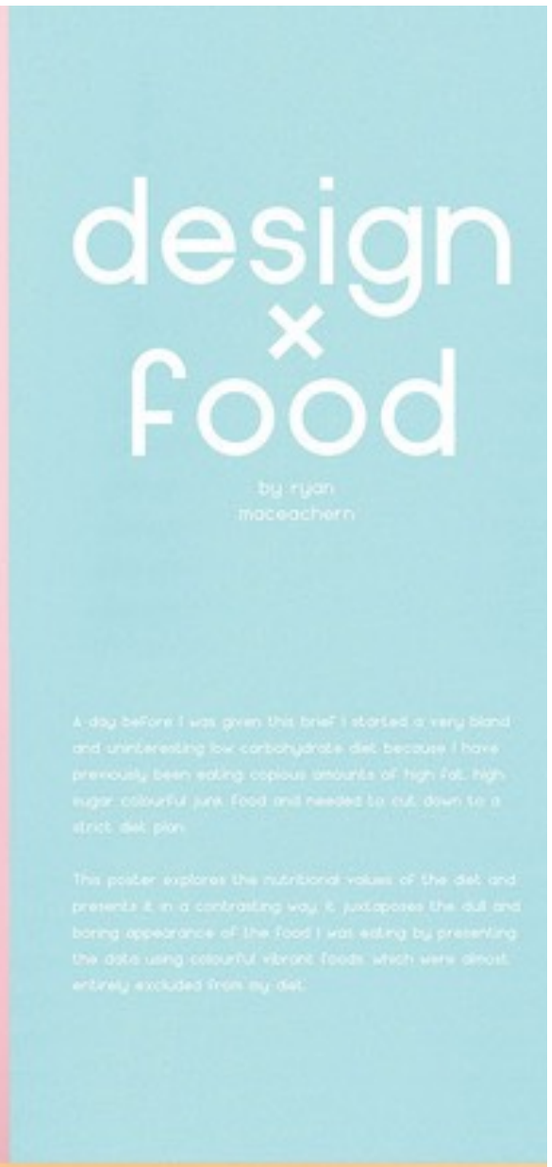
http://mkweb.bcgsc.ca/deadlygenomes/

- Use
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*


Enjoy

Infographics, art, or superfluous visualizations.





https://www.informationisbeautifulawards.com/showcase/67-design-x-food

- Use
  - Consume
    - *Discover*

https://www.behance.net/gallery/924345/Information-graphics-in-context

How can you encode information optimally?

Encode

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, ...

Manipulate

- Change
- Select
- Navigate

Facet

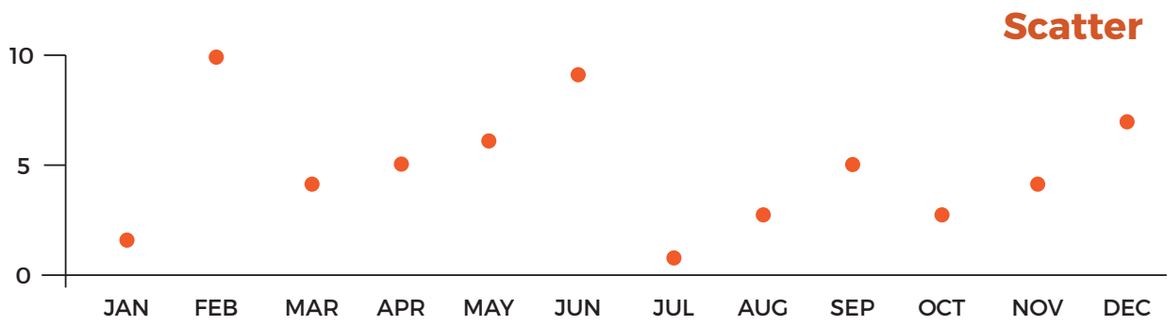
- Juxtapose
- Partition
- Superimpose

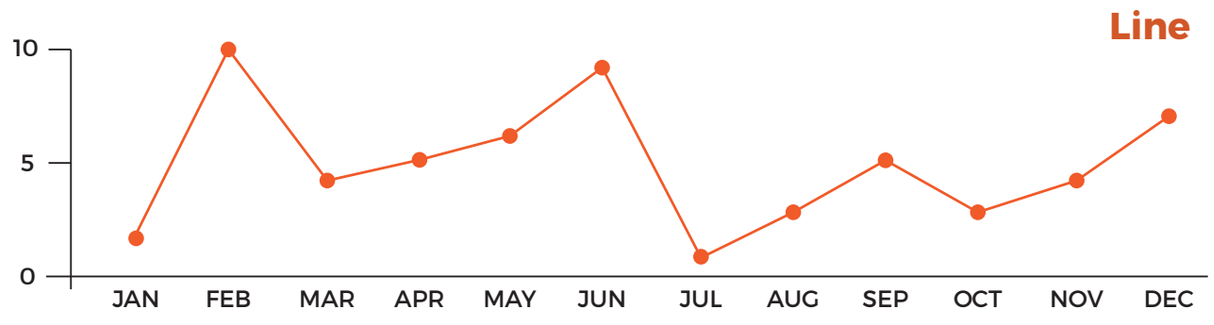
Reduce

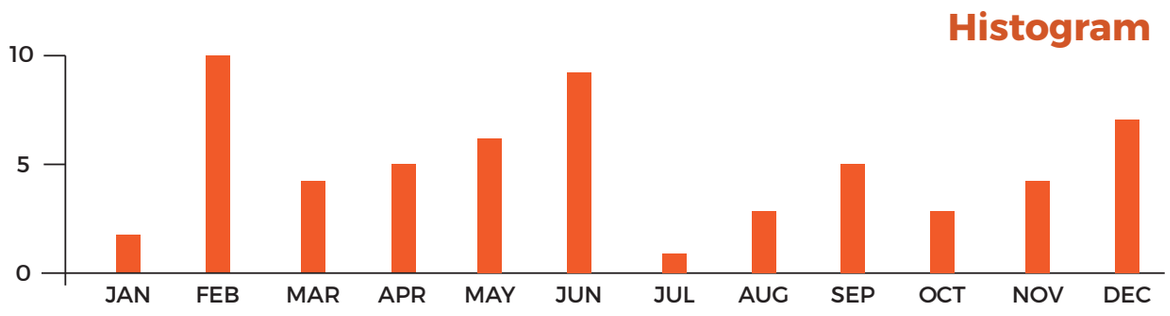
- Filter
- Aggregate
- Embed

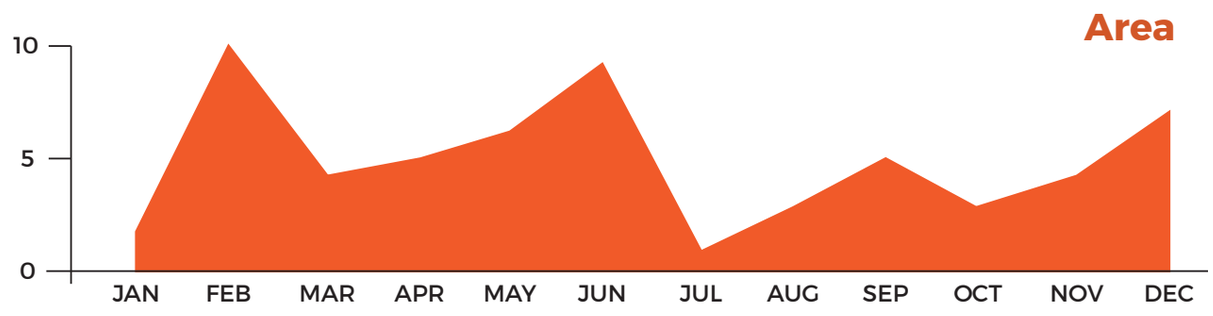
How can you encode information optimally?

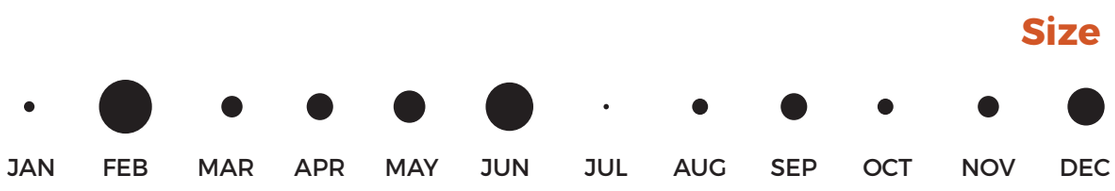
JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	10	4	5	6	9	1	3	5	3	4	7

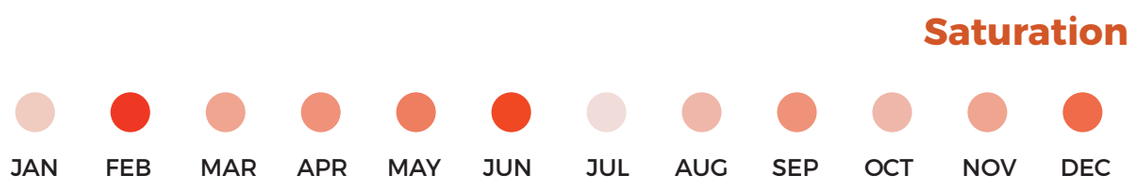












Graphs are like jokes.
If you have to explain them, they didn't work.

Anon.

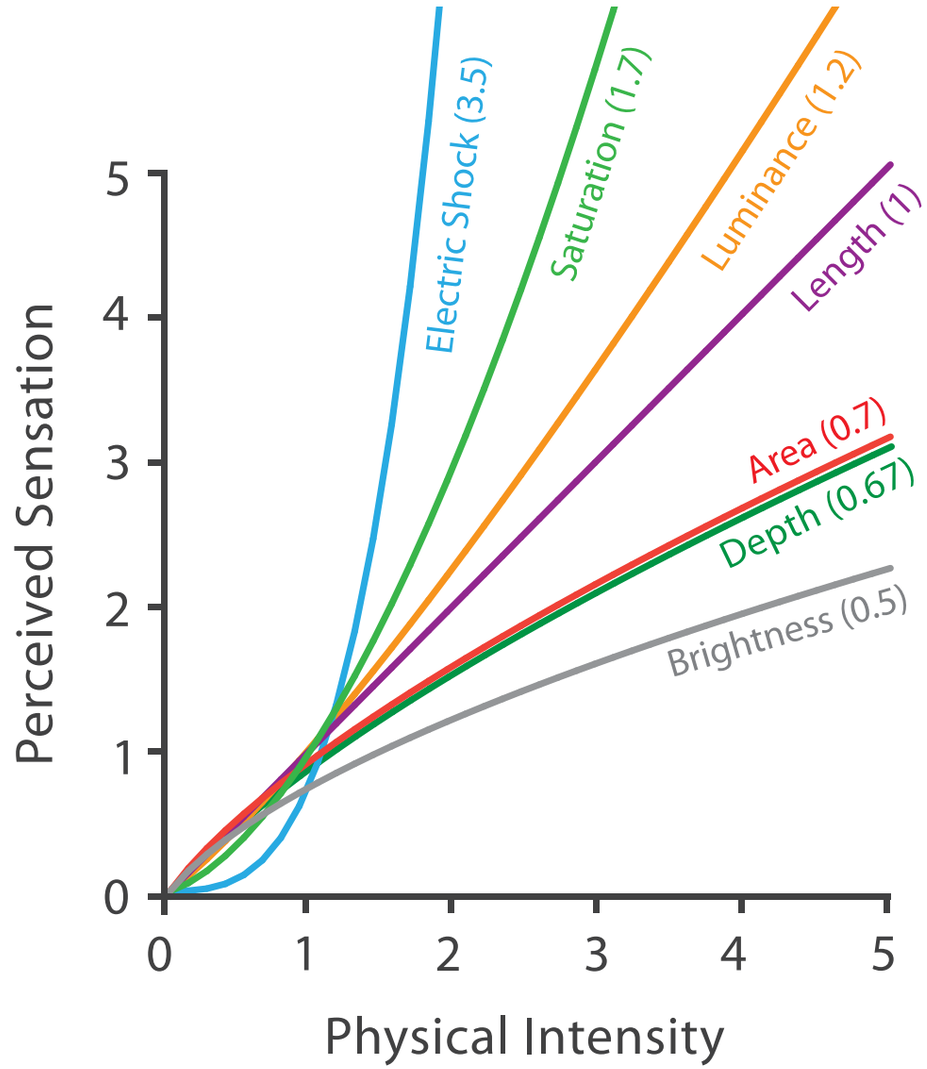
And that's just a really simple low dimensional example

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

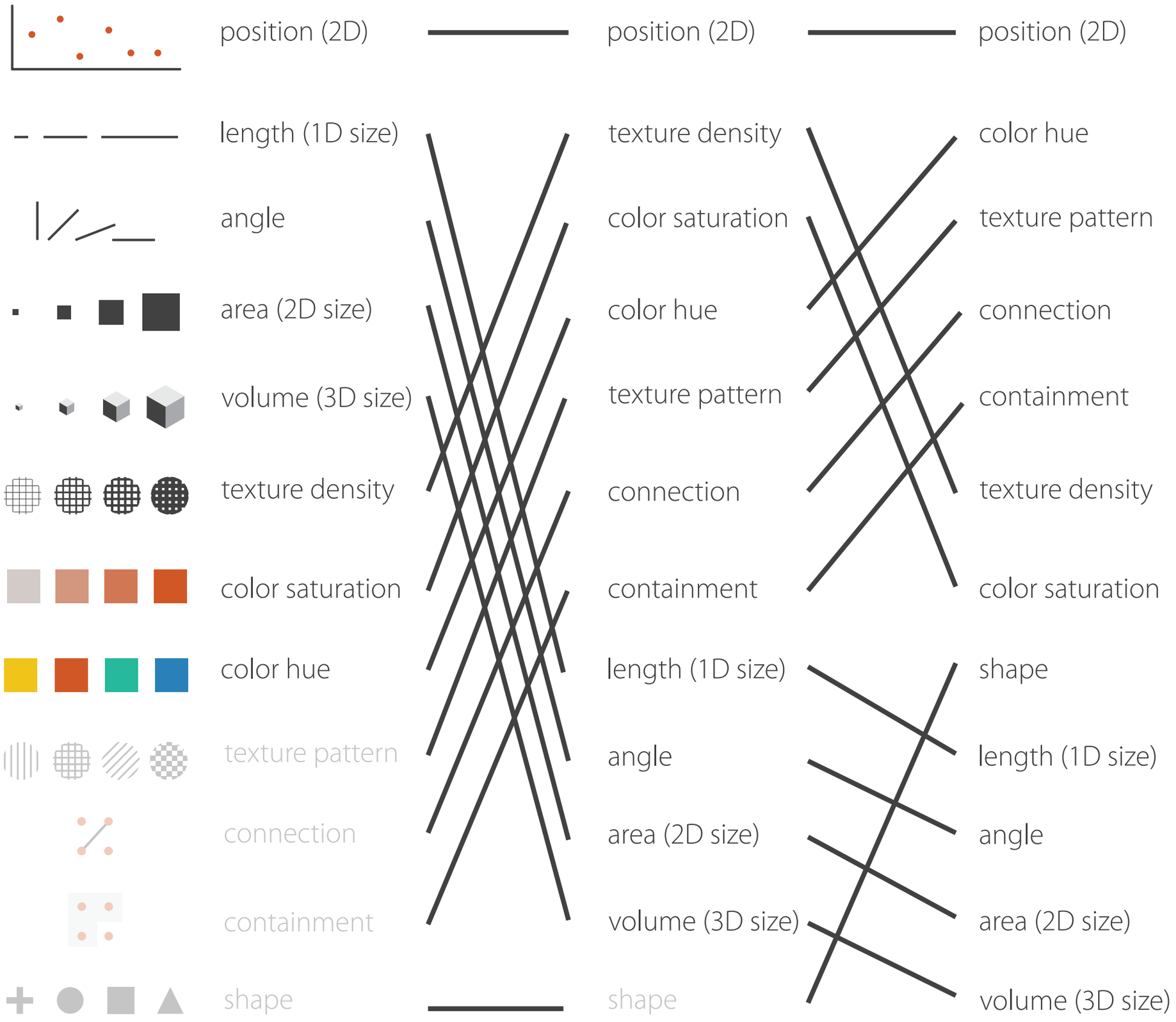
Ordinal not validated

MacKinley, 1986

Categorical not validated

MacKinley, 1986

Suitability of Channel



Positions

T1

T2

T3

T4

T5

Angles

T6

Circular Areas

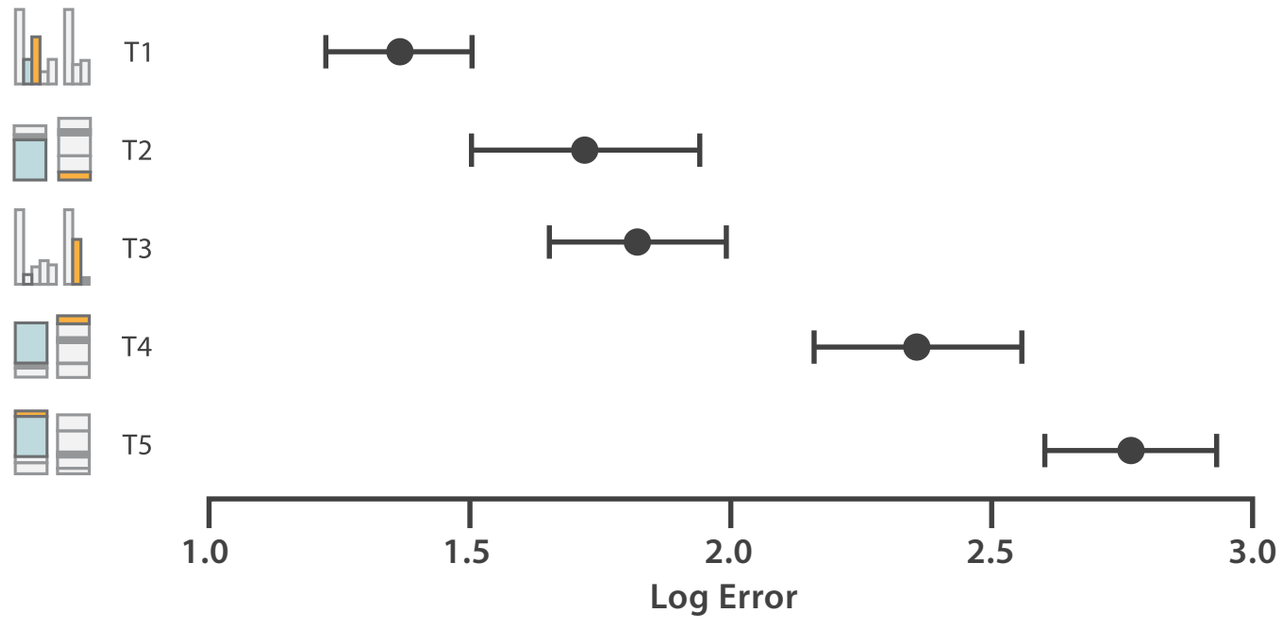
T7

Rectangular areas (aligned or in a treemap)

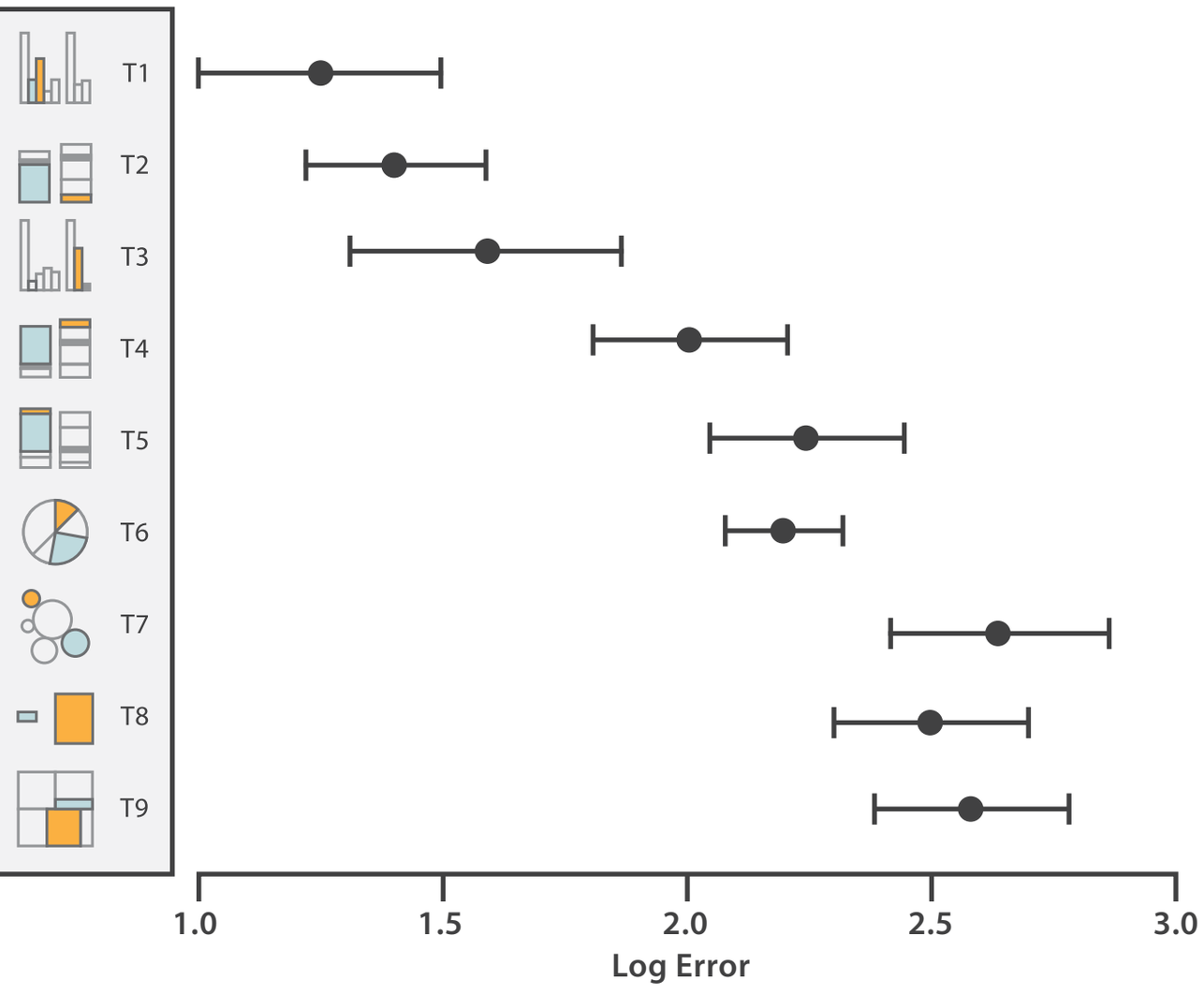
T8

T9

Cleveland & McGill's Results 1984

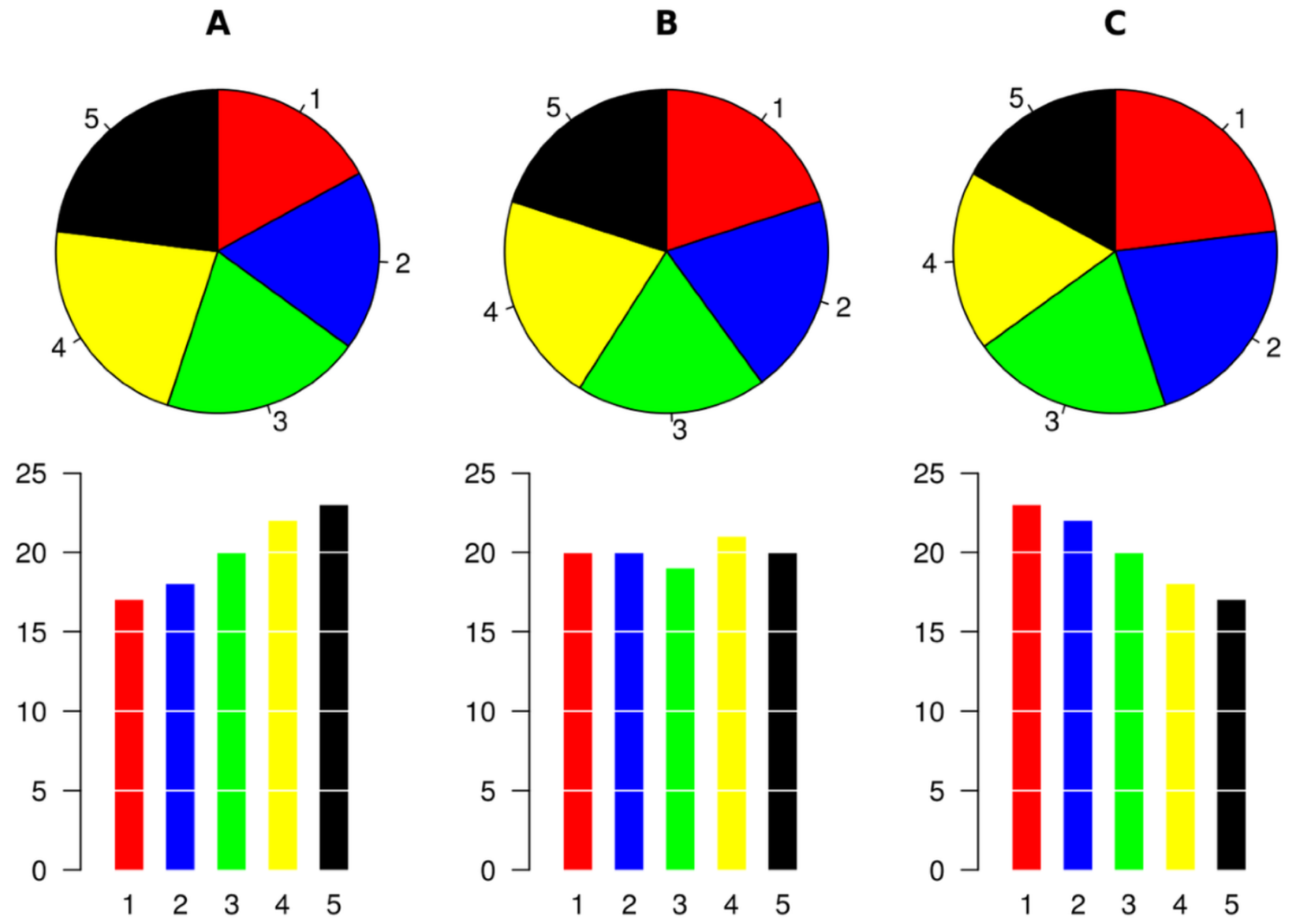


Heer and Bostock 2010 Crowdsourced Results



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

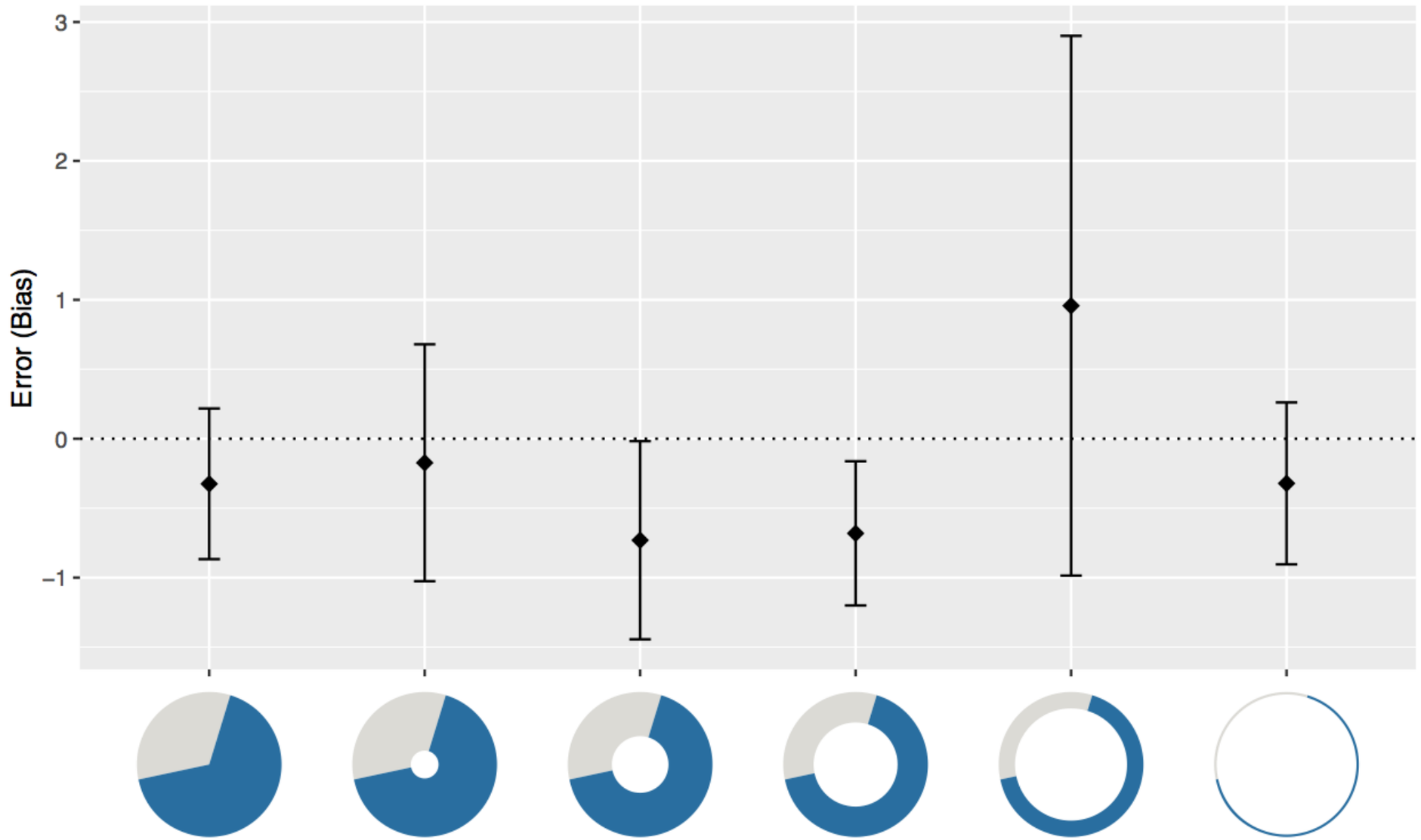
When someone reads or compares values in a pie chart, what are they doing? Comparing angles, areas, length of arc?



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?

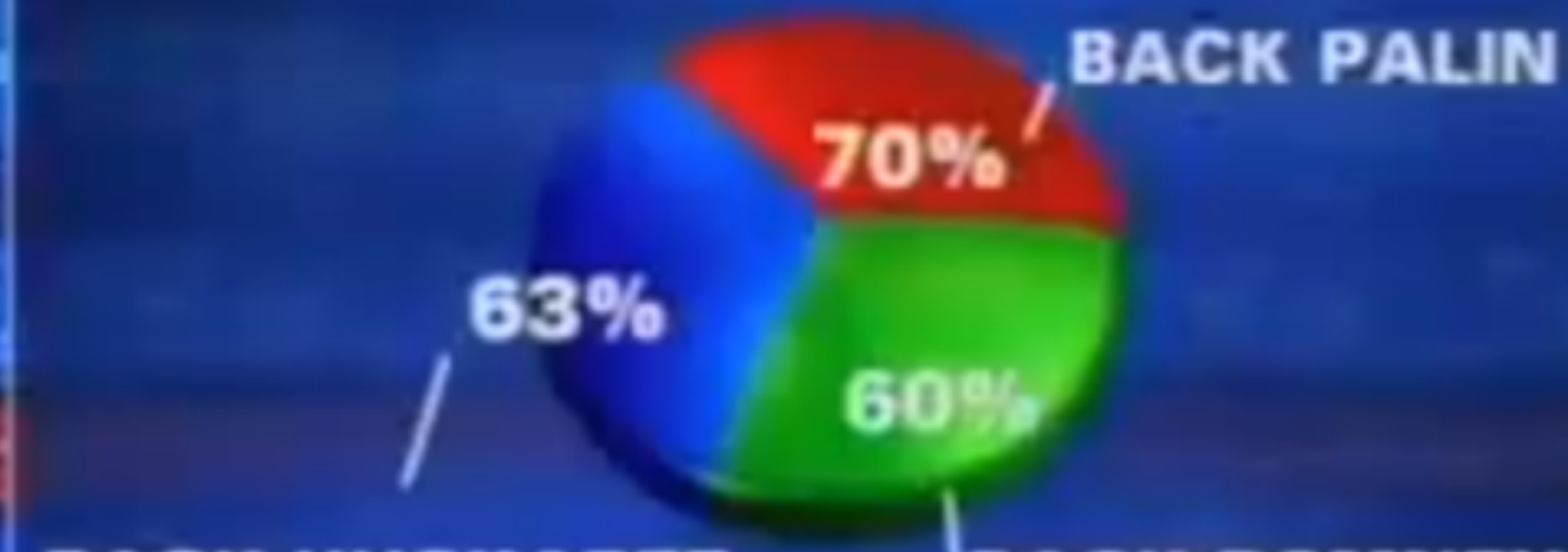


Robert Kosara and Drew Skau. 2016. **Judgment error in pie chart variations**. In Proceedings of the Eurographics: Short Papers (EuroVis '16). Eurographics Association, Goslar Germany, Germany, 91-95. DOI: https://doi.org/10.2312/eurovisshort.20161167

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

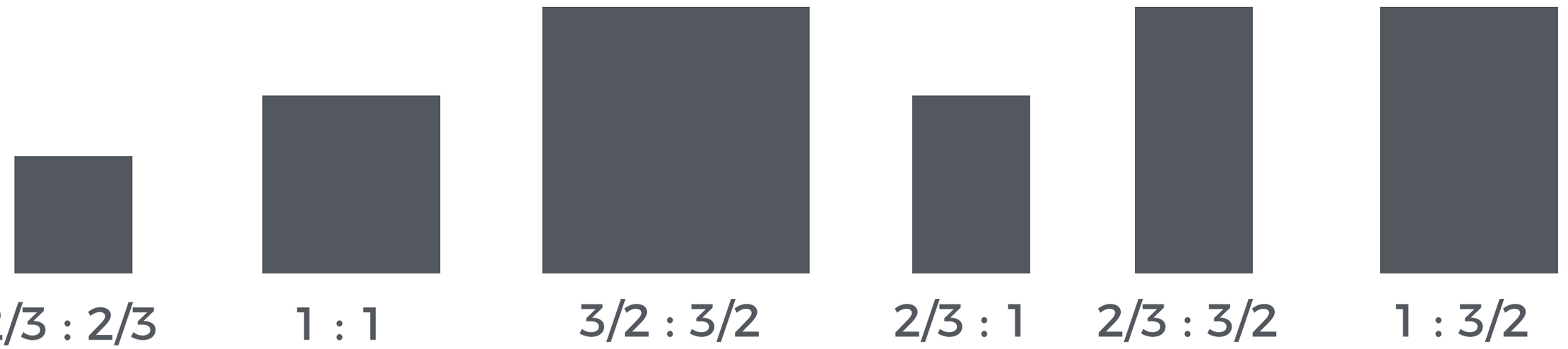


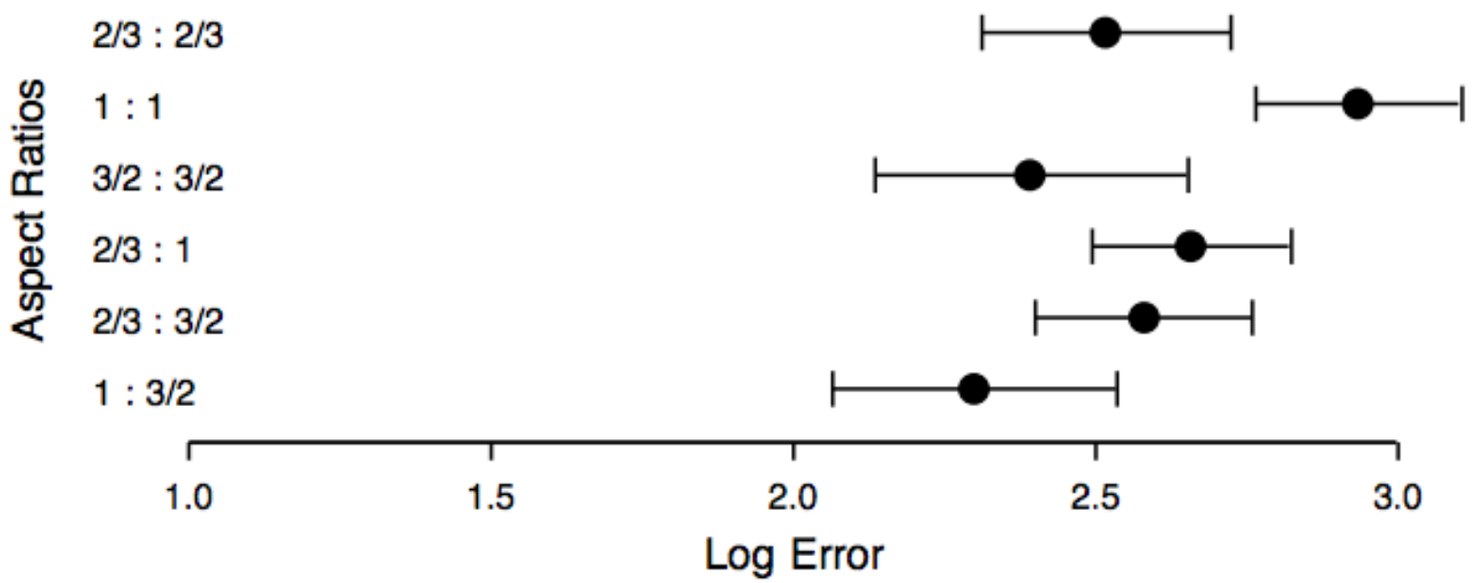
FOX

9:17 PM

SOURCE:OPINIONS DYNAMIC

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





These results are directly relevant treemap construction.

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.

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

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

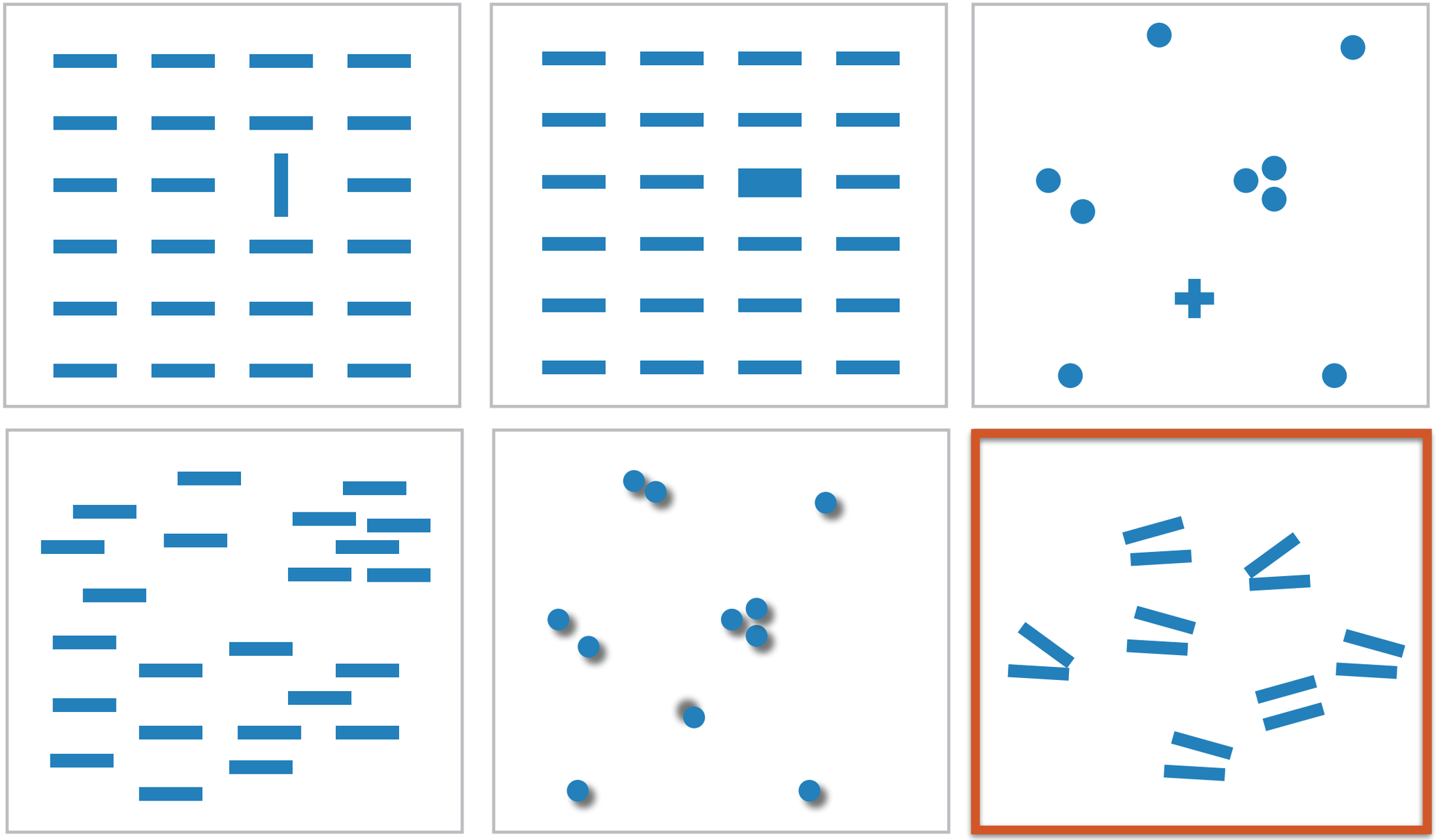
Some data has a **natural mapping** that our brains expect given certain types of data

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



Not all exhibit the pop-out effect!

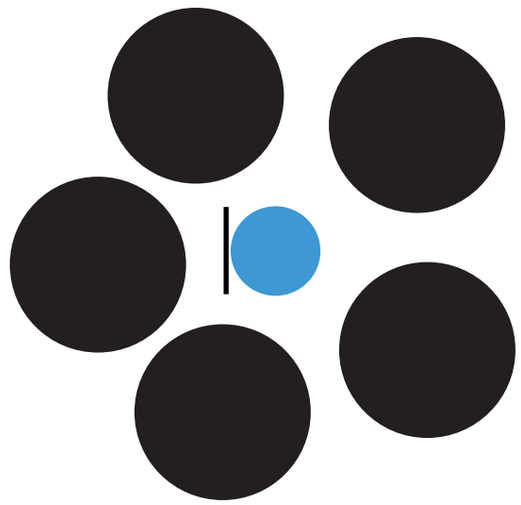
Parallel line pairs do not pop out from tilted pairs...

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

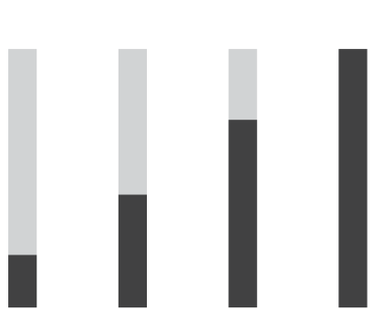
Relative Comparison

Relative Comparison

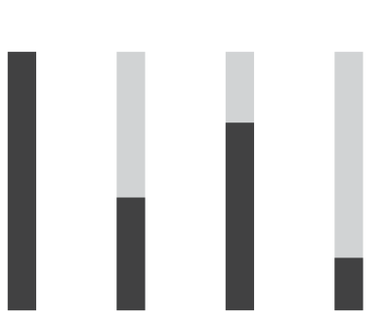
36px |



Relative Comparison

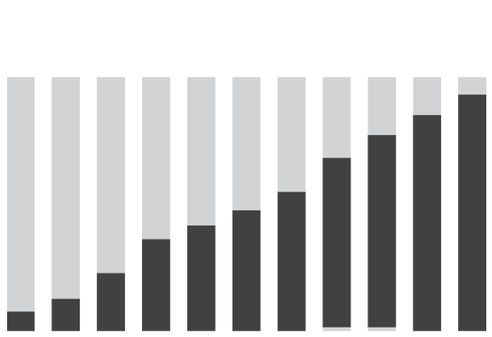


4 values

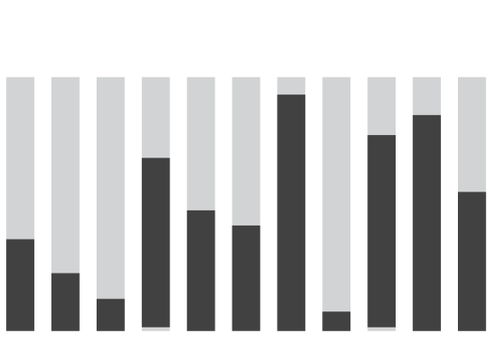


Unordered

Unaligned



11 values



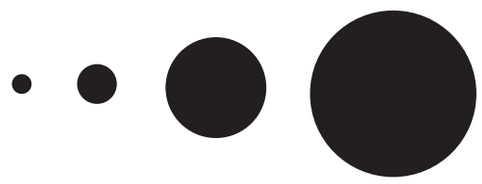
Unordered

Unaligned

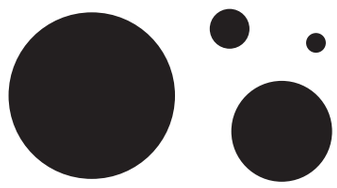
Relative Comparison

4 values

Aligned

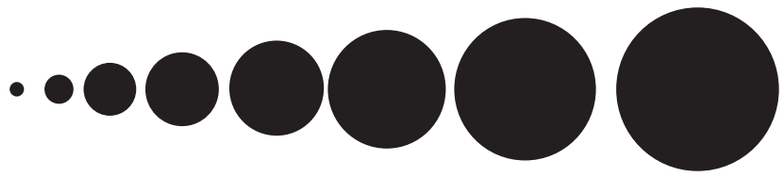


Unordered

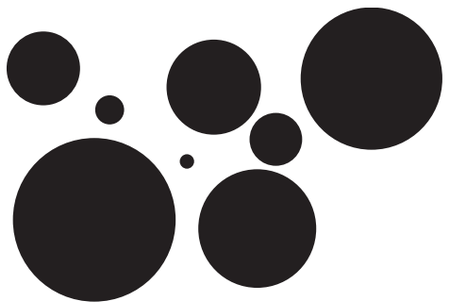


8 values

Aligned



Unordered



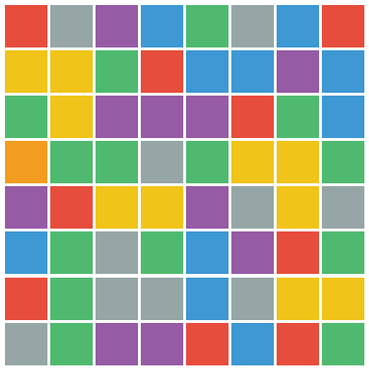
Relative Comparison

8 values

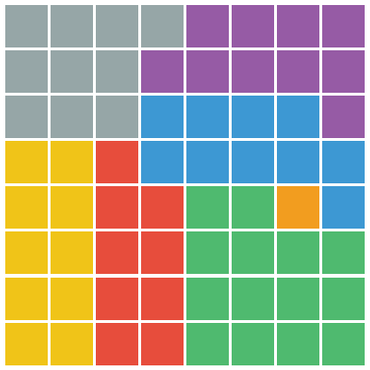
20 values

A) Known and Unknown Target Search

Random



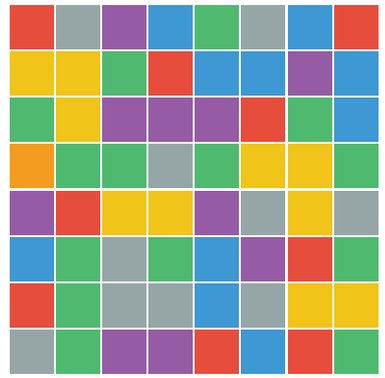
Grouped



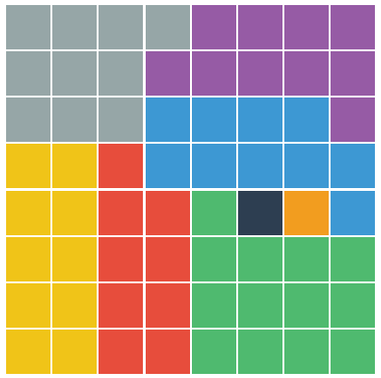
Target shown before hand (**known**) or **not shown** (**unknown**).
The unique colour here is the orange square.

B) Subitizing (how many colours?)

Random

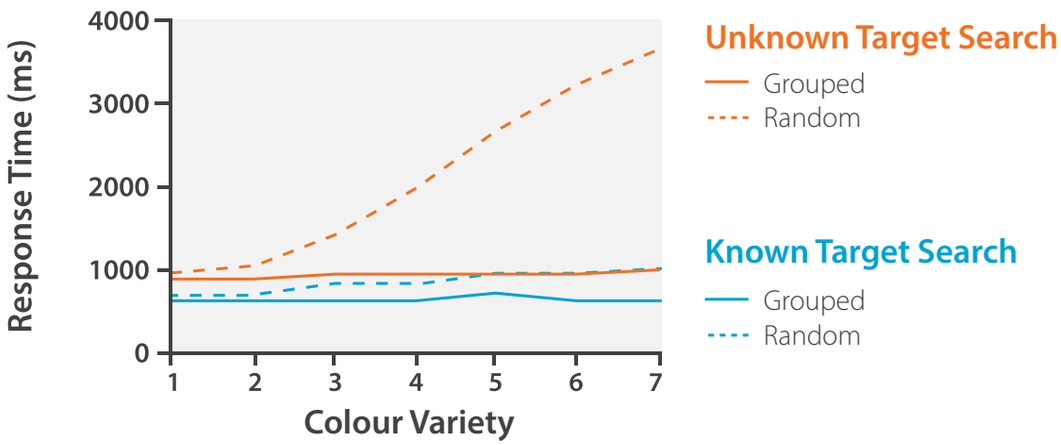


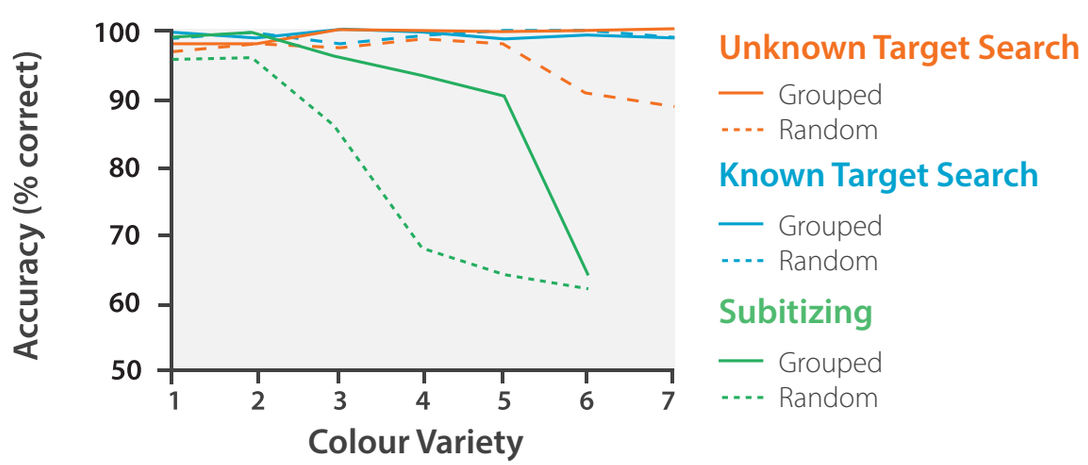
Grouped



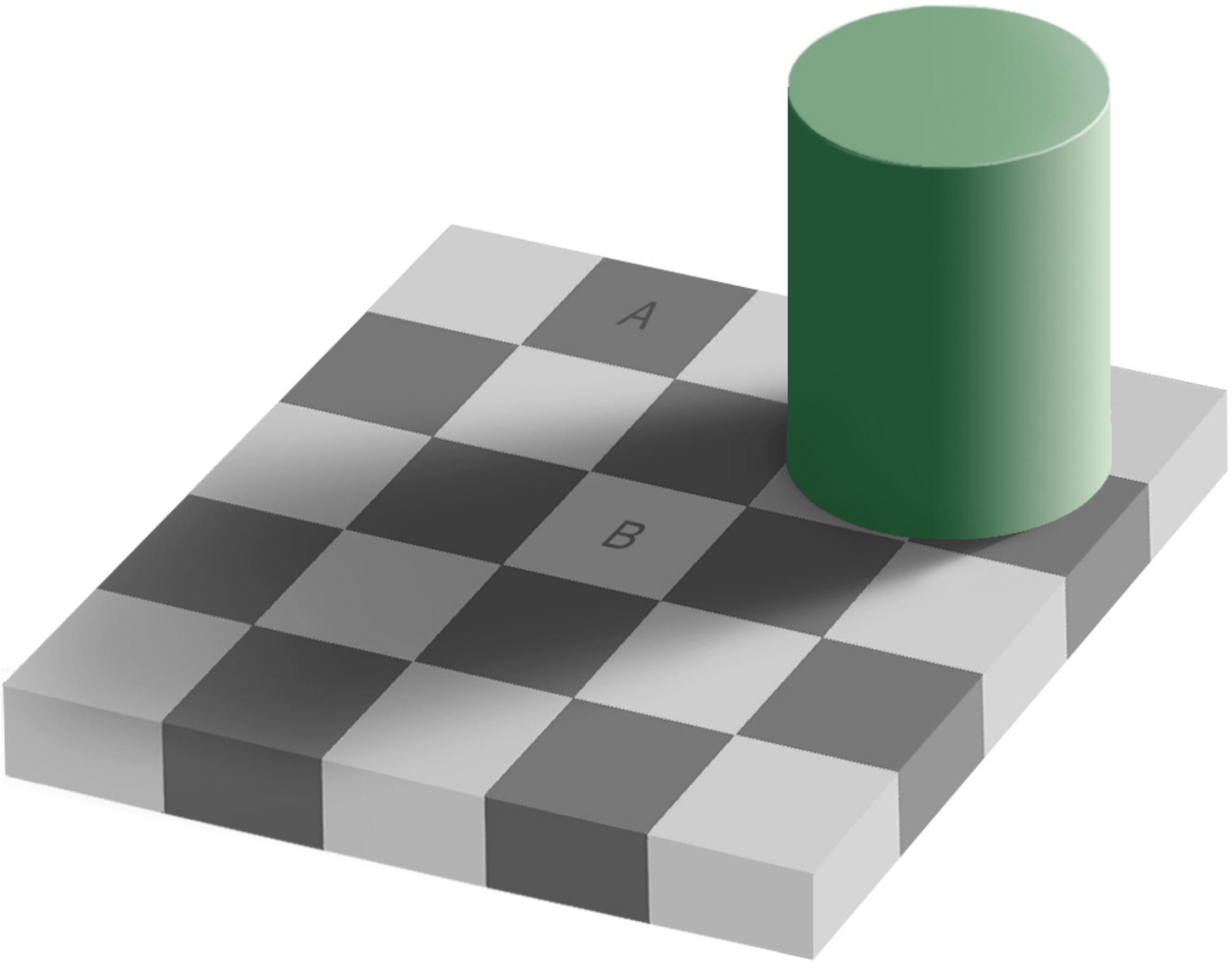
Which grid has more colours?

C) Response Time and Accuracy Results





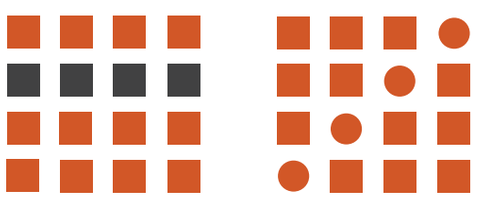
How Capacity Limits of Attention Influence Information Visualization Effectiveness.
Haroz S. and Whitney D., IEEE TVCG 2012


A
B

Gestalt Laws

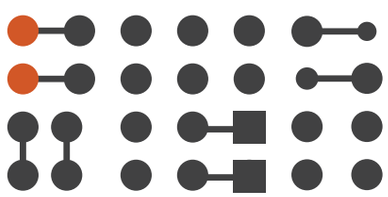
A. Law of Closure

B. Law of Similarity



C. Law of Proximity

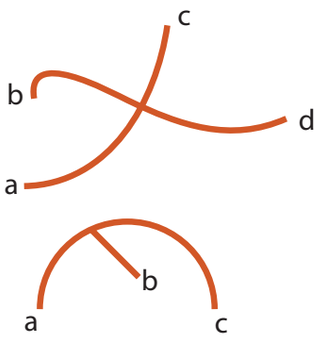
D. Law of Connectedness



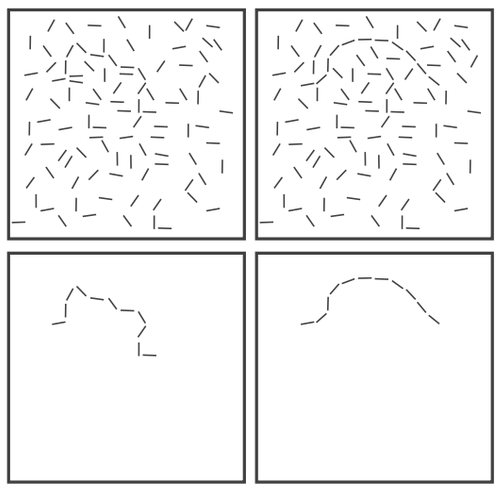
E. Law of Symmetry

[ ] { } ( )

F. Law of Good Continuation



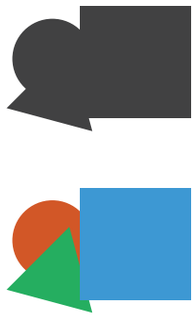
G. Contour Saliency



H. Law of Common Fate

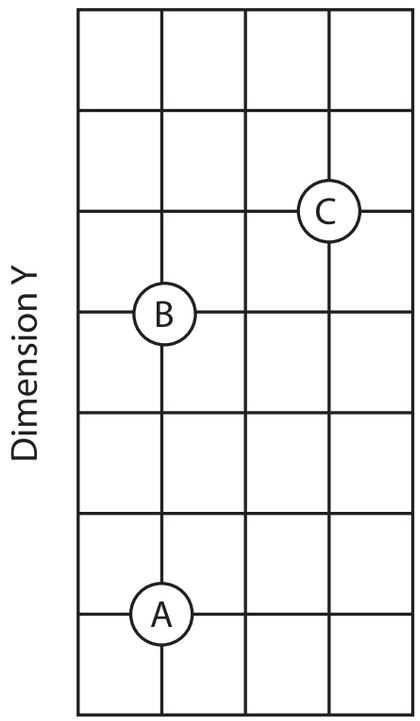
I. Law of Past Experience

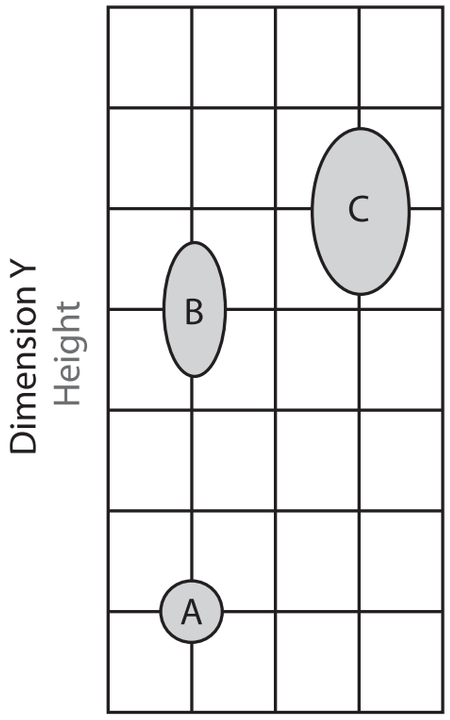
J. Law of Pragnanz

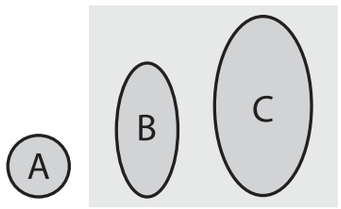


K. Figure/Ground

Integral/Separable Dimensions

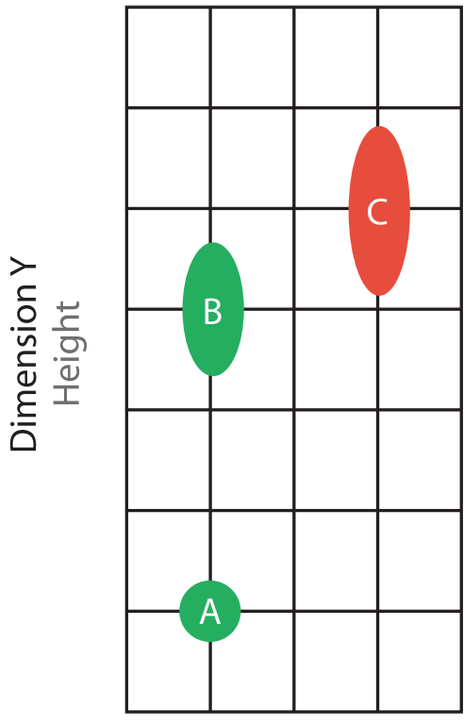


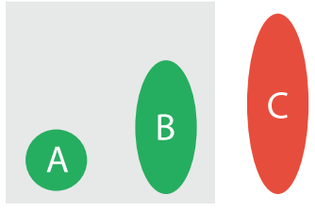




A and **B** have the same width. However **B** and **C** are perceived more alike even though they are different widths and heights.

Width and **Height** are **integral** dimensions

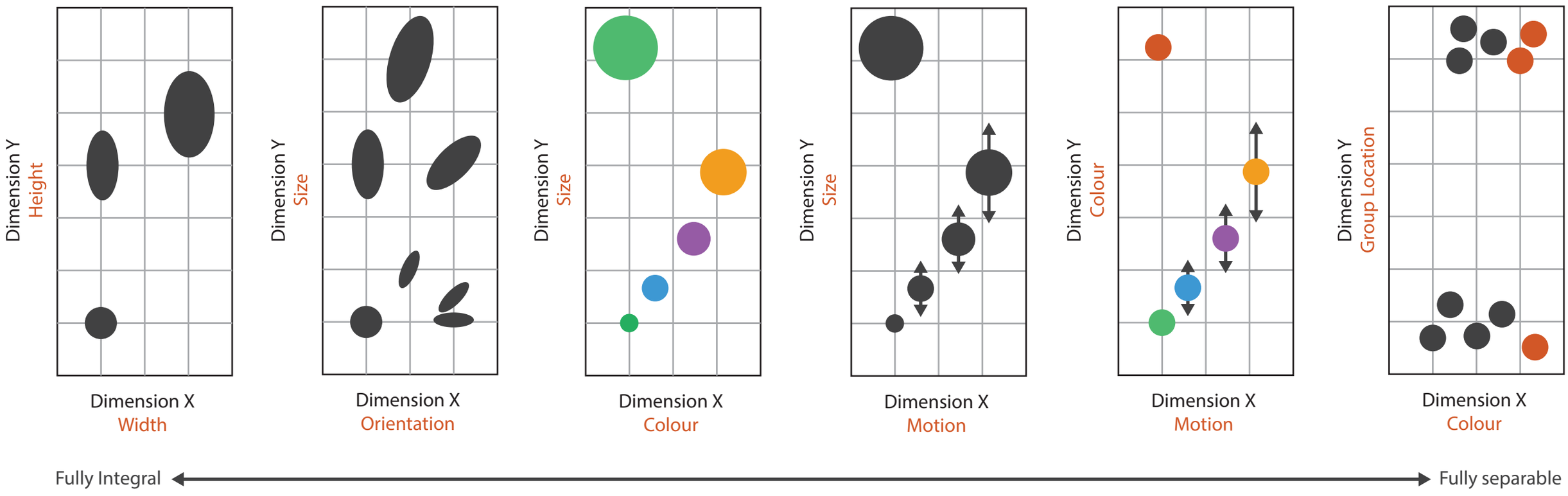




A and **B** have the same colour and are perceived more similar.

Colour and **Height** are **Separable** dimensions

Integral/Separable Dimensions



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

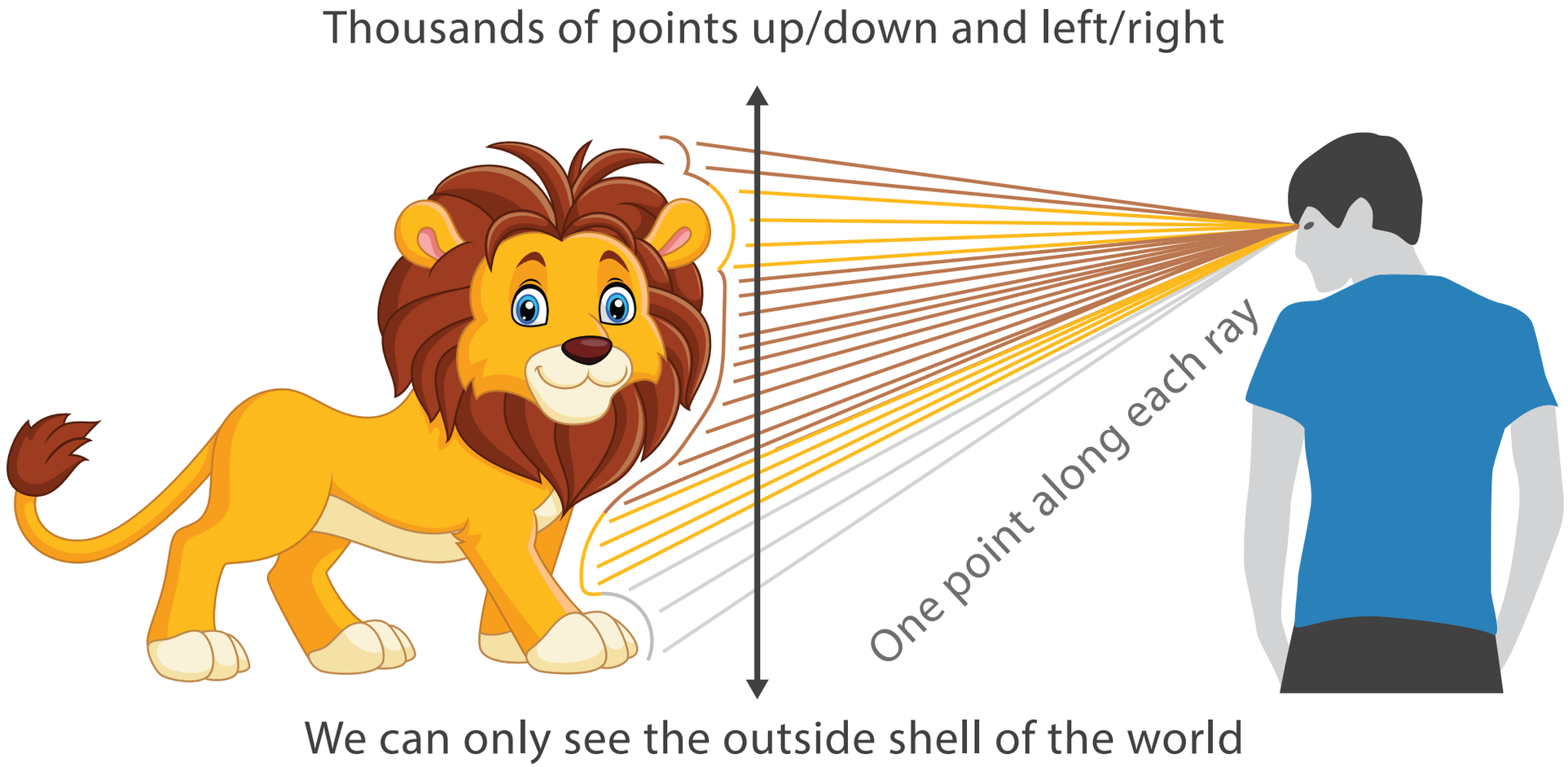
Some visual channels are more effective for some data types over others.

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.

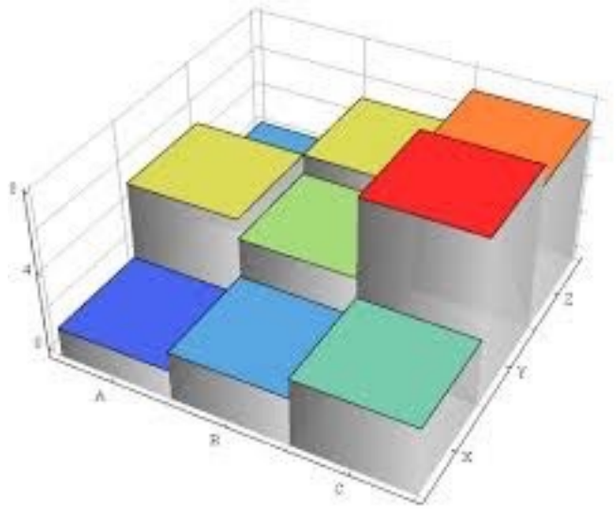
2D always wins...

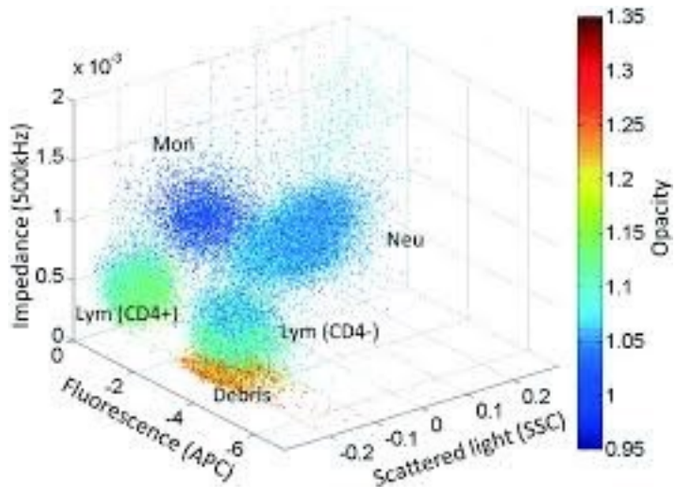


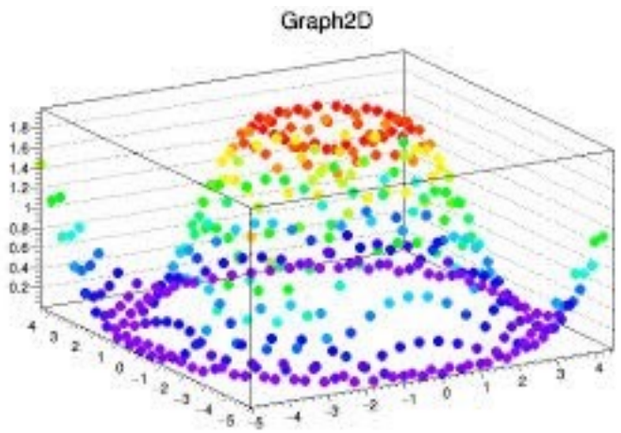
Our visual system is not good at interpreting information on the z-axis.

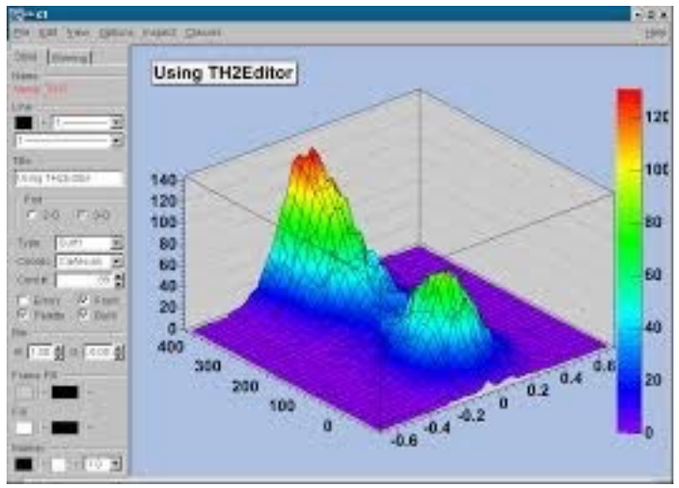
*3D is normally only used for exploration of inherently 3D information, such as medical imaging data…

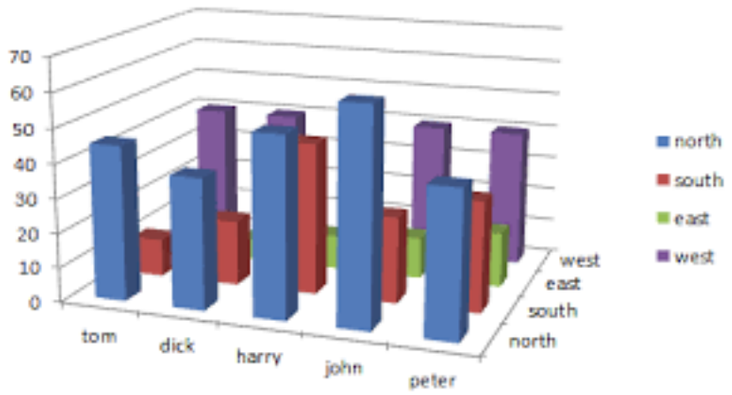
2D always wins...





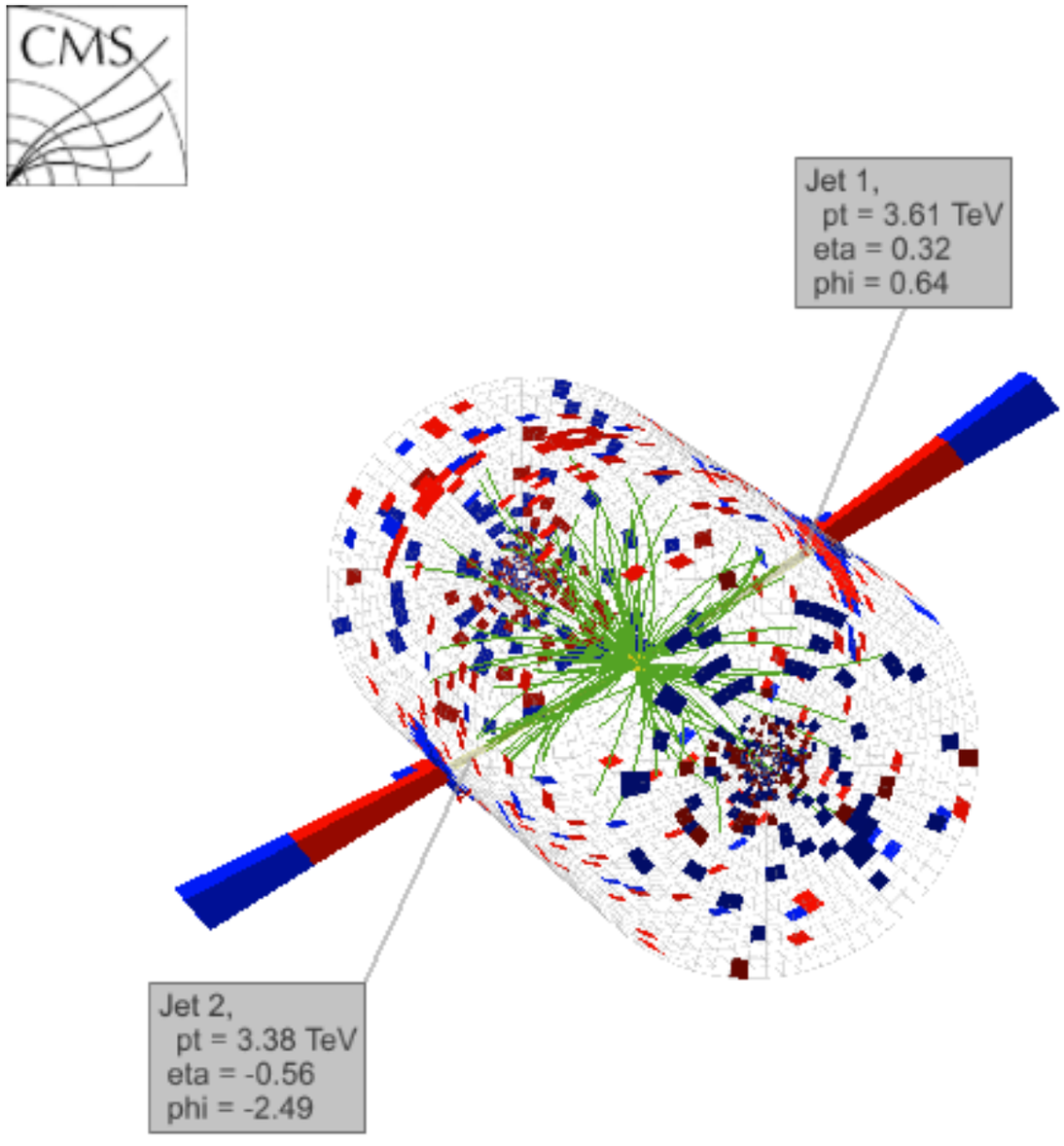


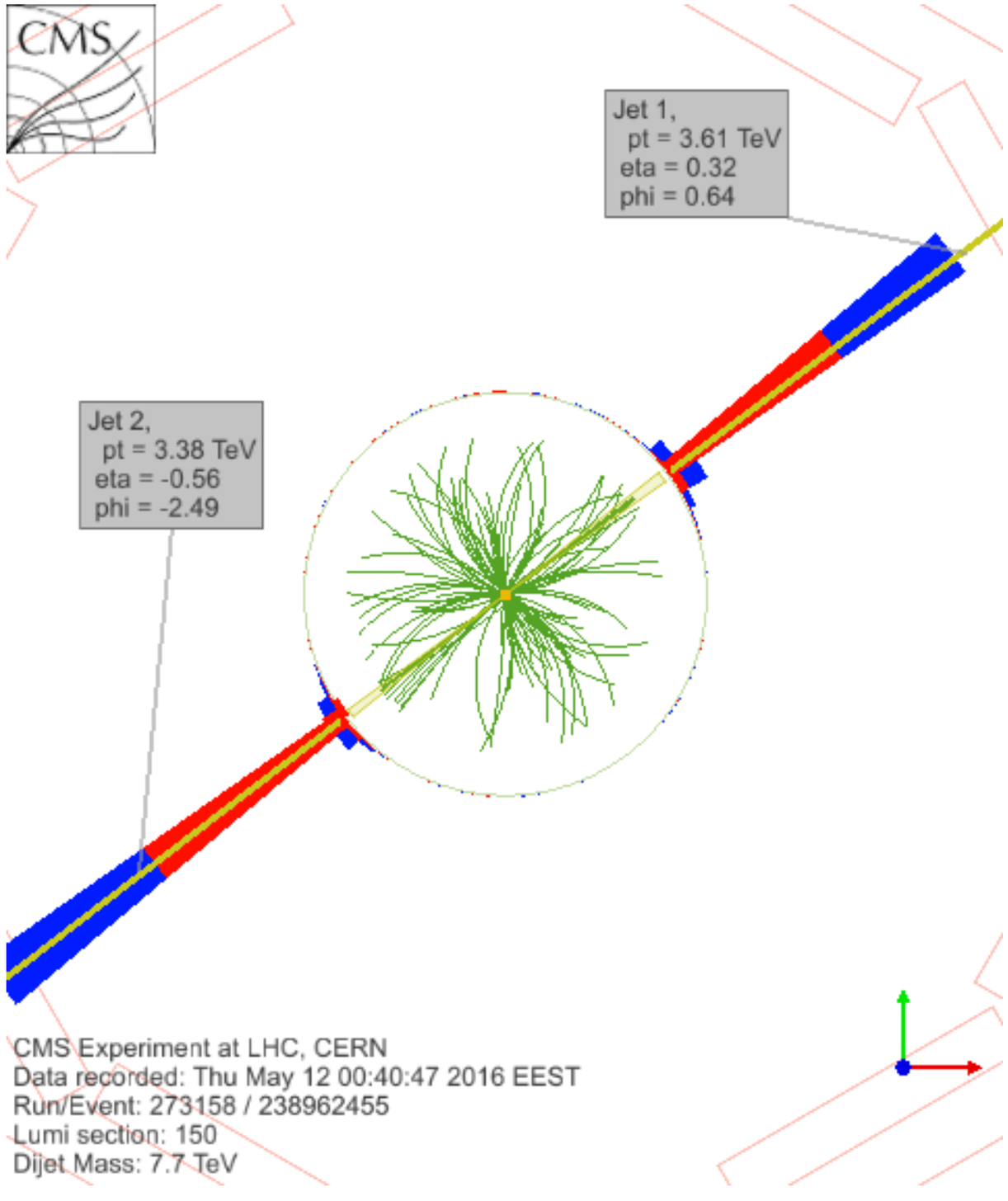




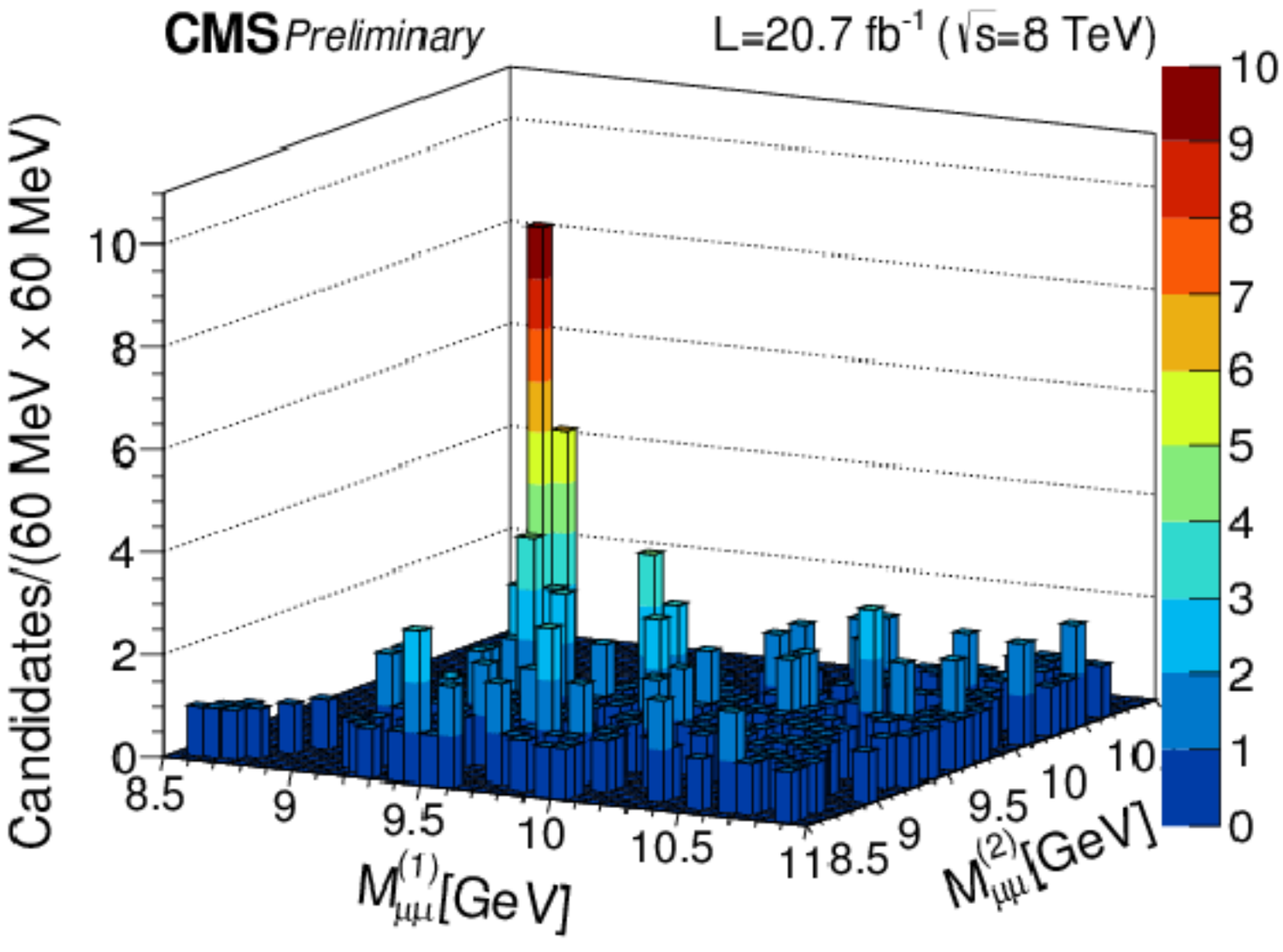
These options, taken randomly from google image searches so how widely 3D is abused in information visualization. All of these charts are manipulating our perception of the data by using the Z axis to occlude information…it would be avoided in 2D.

2D always wins...





2D always wins...



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