part one!

biological basis of information design

what visualisations can do for us

purpose - what is your (specific) goal?

data - what kind of data do you have?

visual dimensions - representing data visually

communication - deception and bad infographics

introduction to information graphics & data visualisation

4:3

mieke van heesch (@emax)

University of Southampton *for* Open Data Institute Short Course

0:28 EST

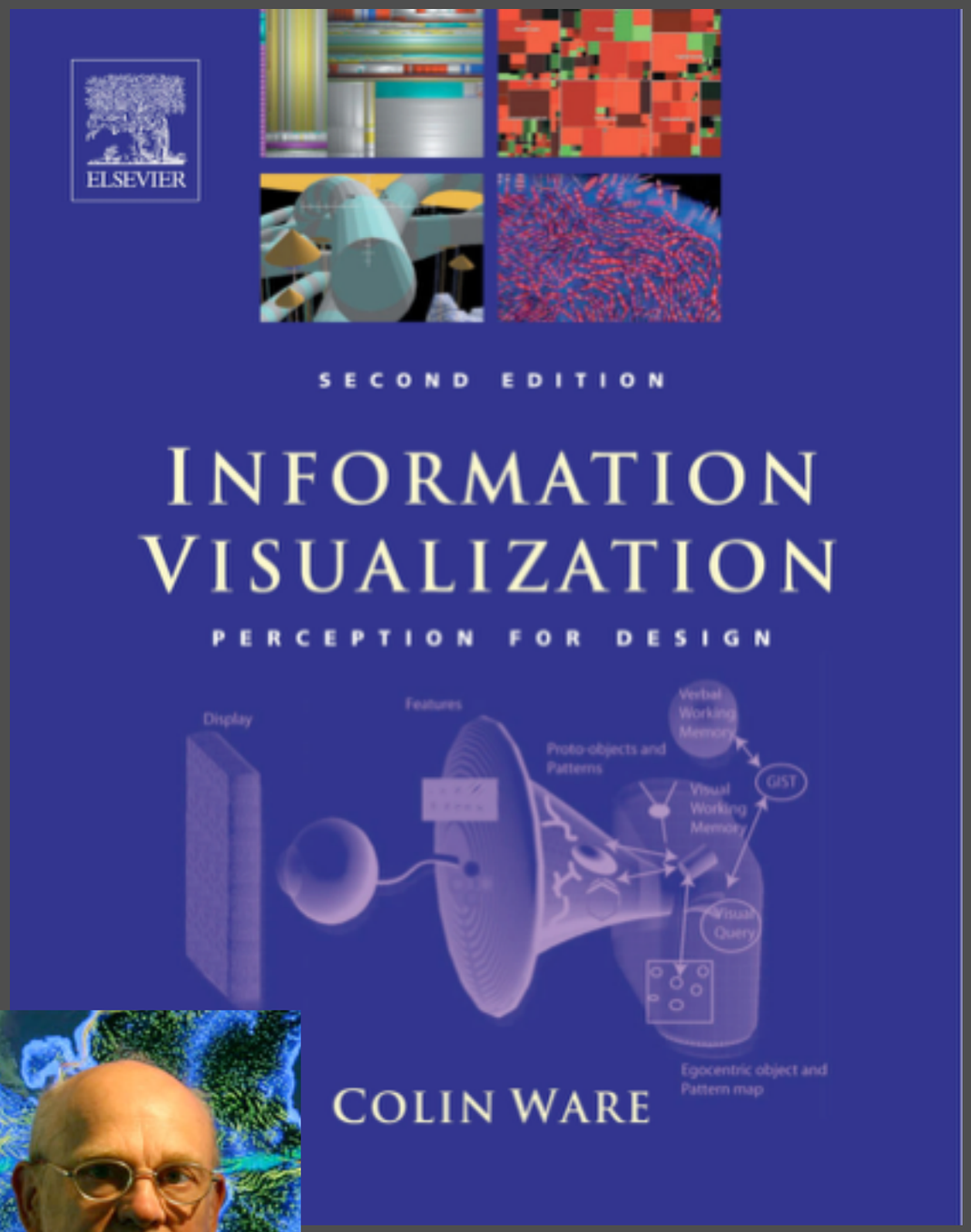
what are the **goals** of visualisation?

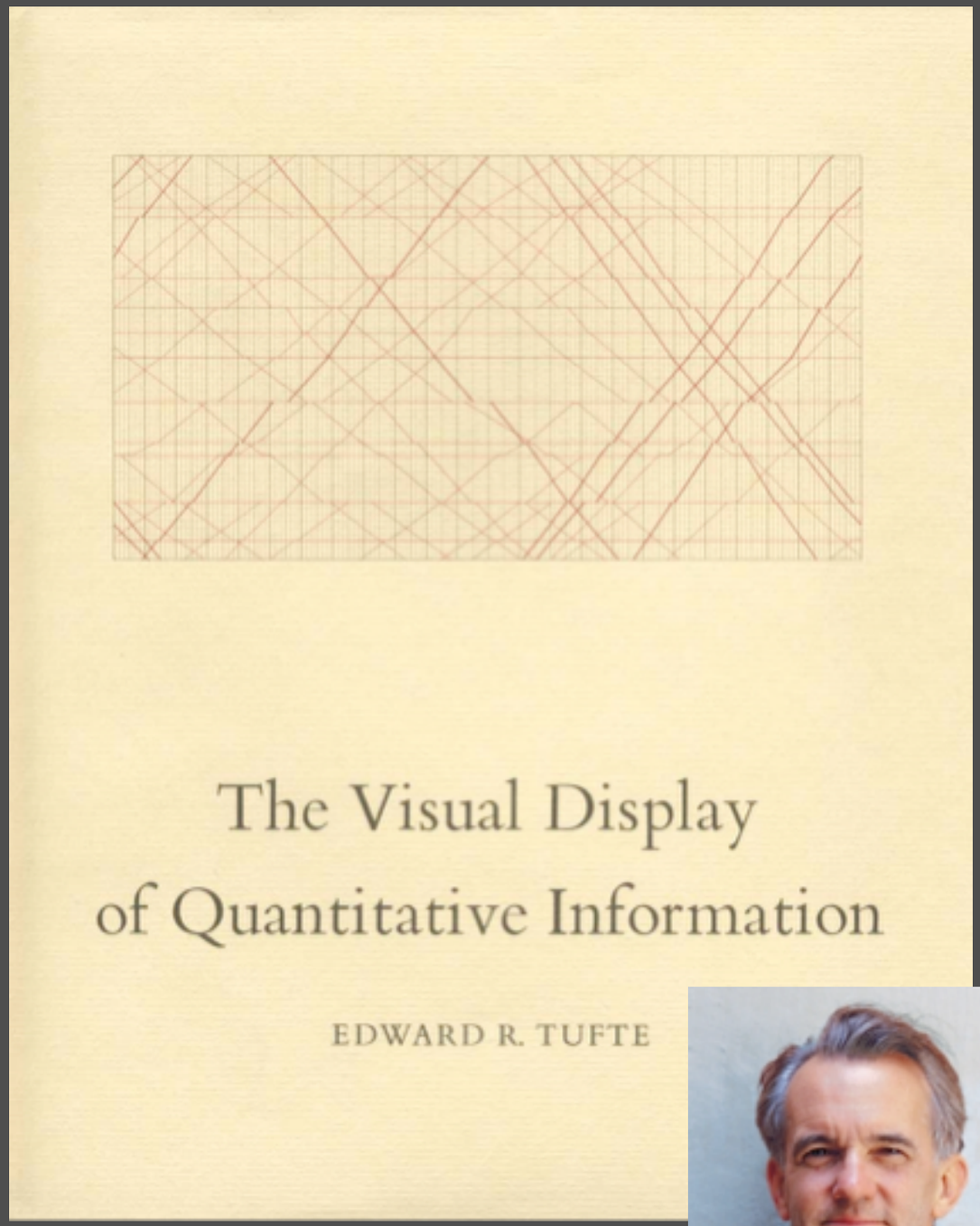
how do you choose a visual representation for data?

how do you evaluate a visualisation?

key objectives

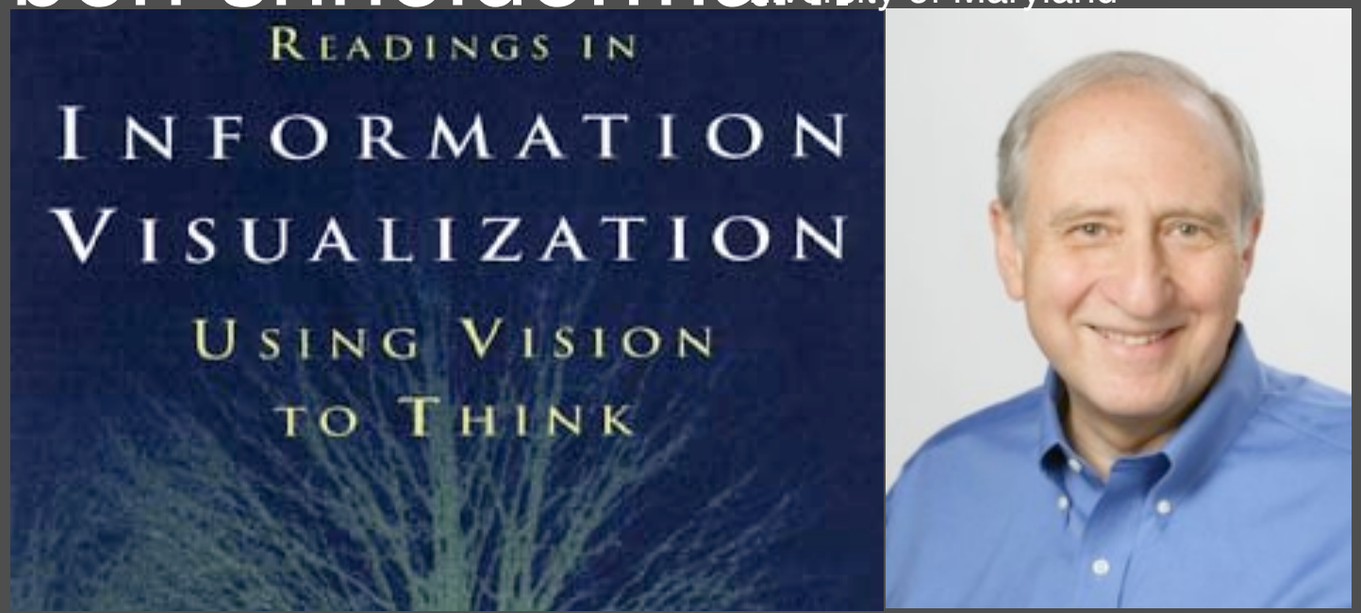
theory


ELSEVIER
SECOND EDITION
INFORMATION VISUALIZATION
PERCEPTION FOR DESIGN
Display
Features
Proto-objects and Patterns
Verbal Working Memory
GIST
Visual Working Memory
Visual Query
Egocentric object and Pattern map
COLIN WARE

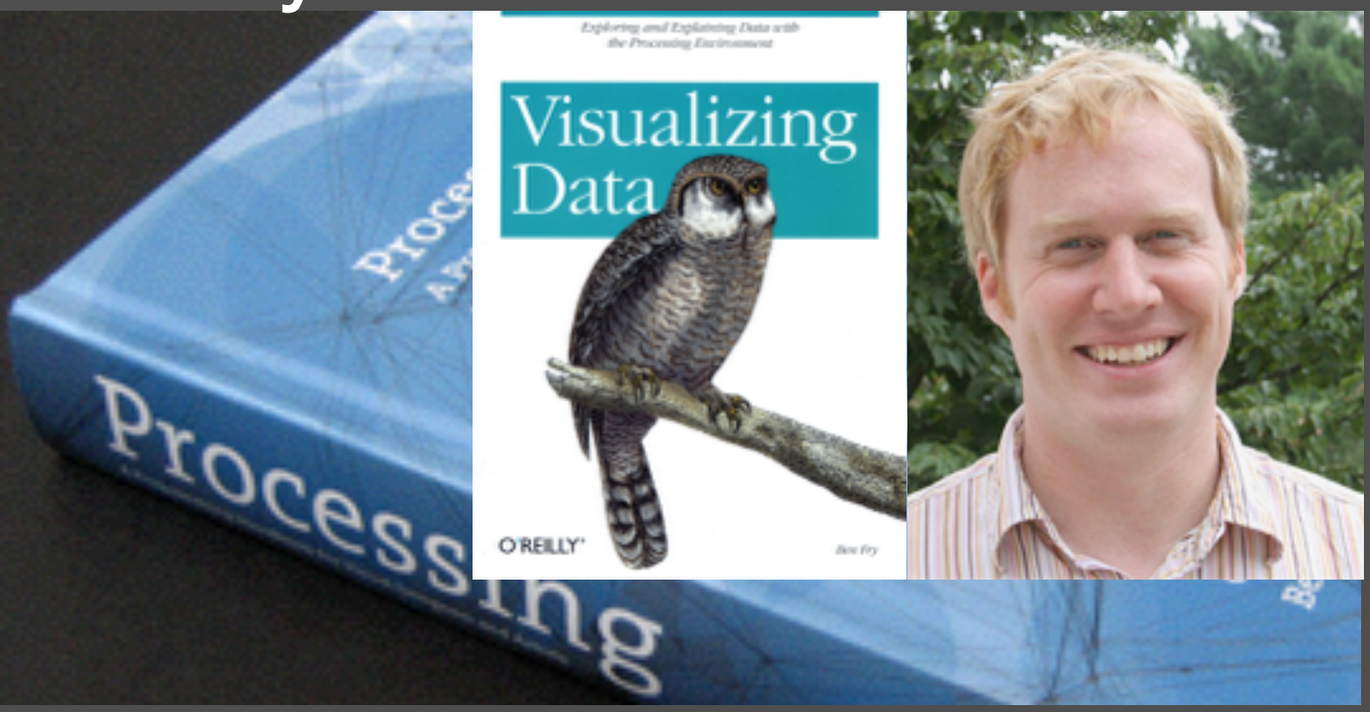

The Visual Display
of Quantitative Information
EDWARD R. TUFTE

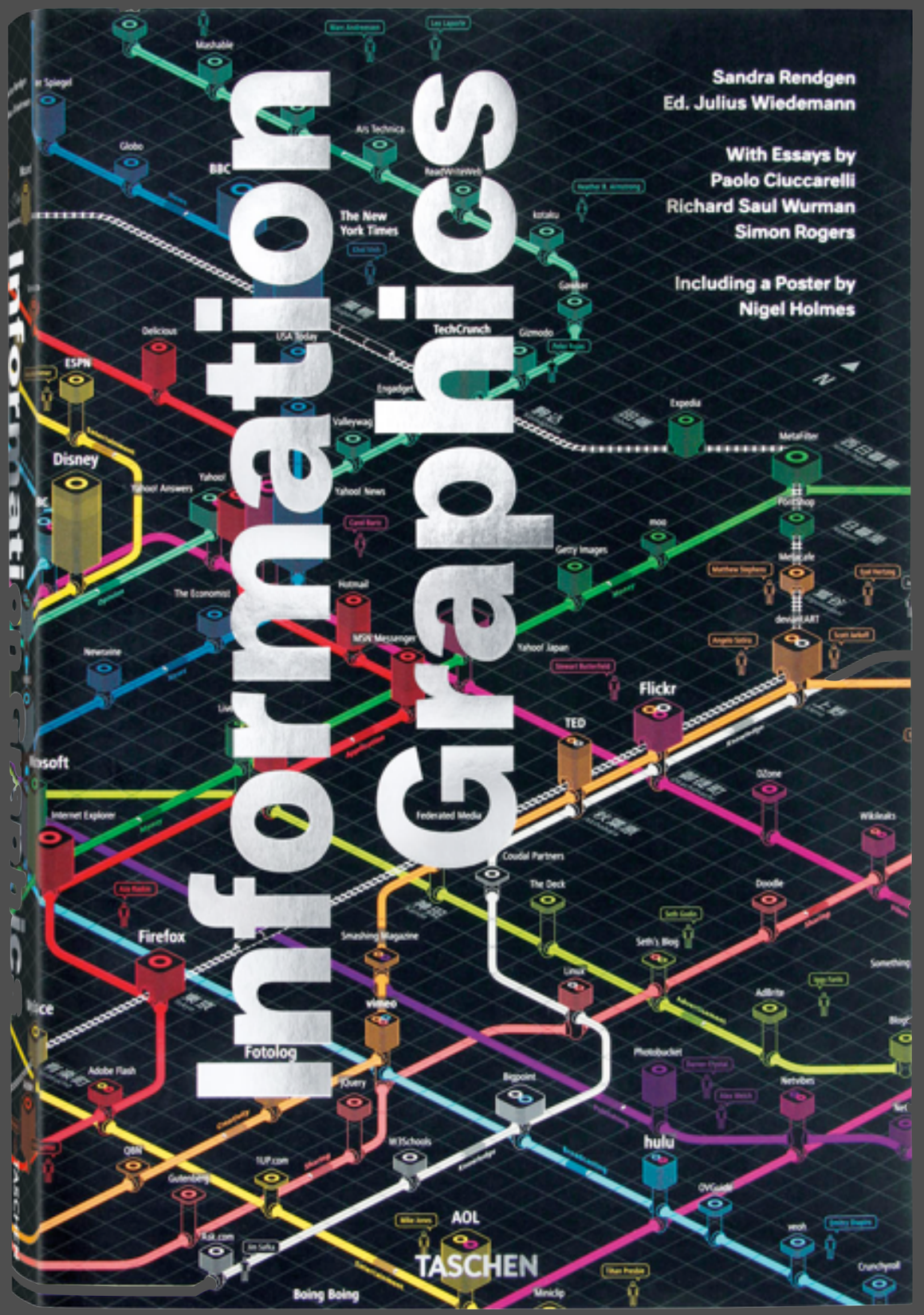
praxis

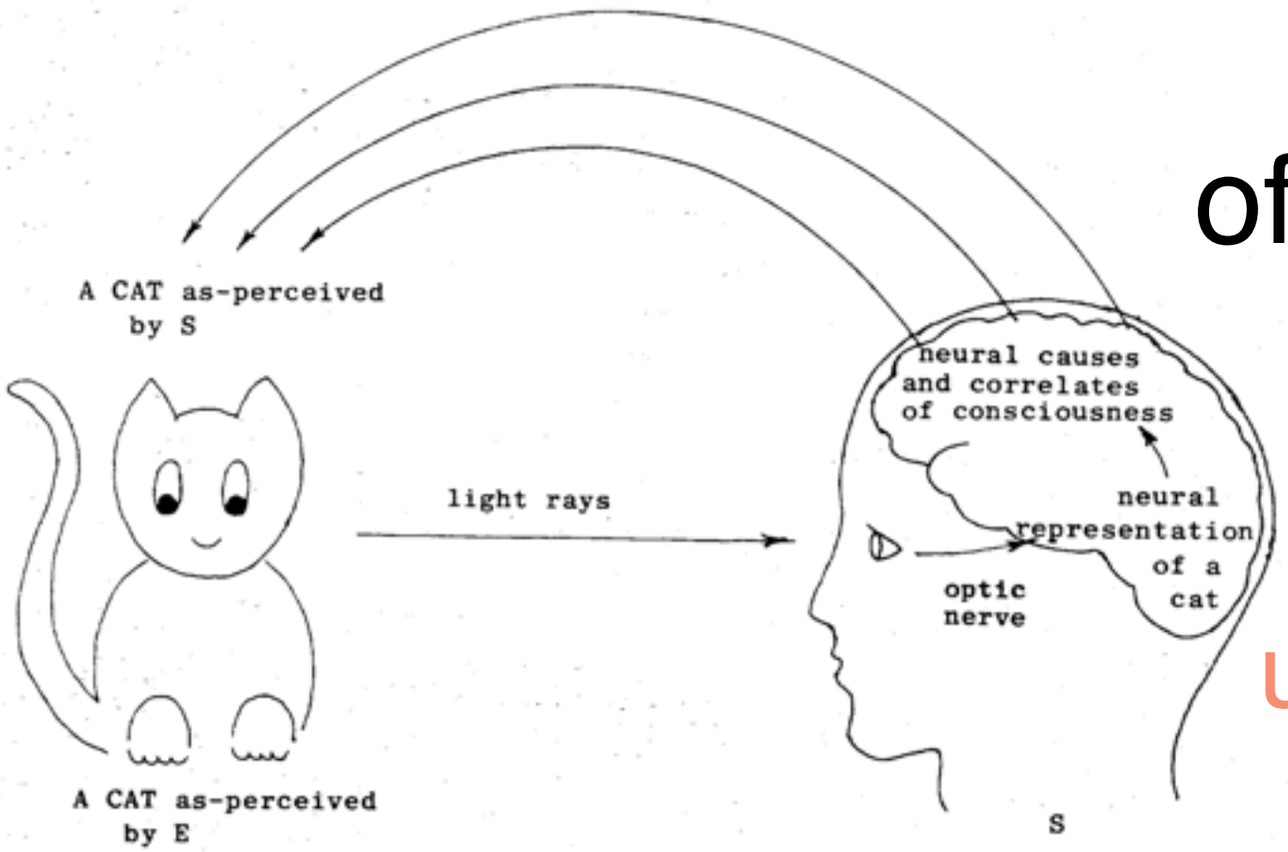
ben shneiderman, University of Maryland



ben fry, MIT Media Lab/fathom.info



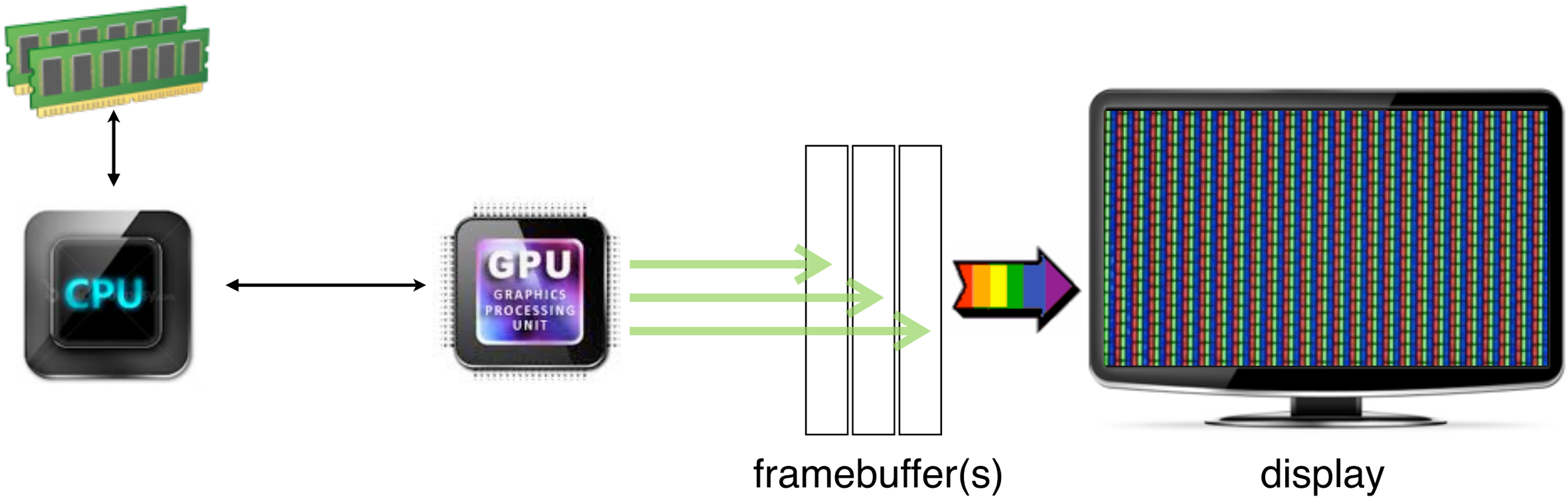




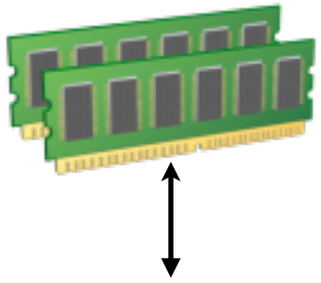
what is the goal of of information design?

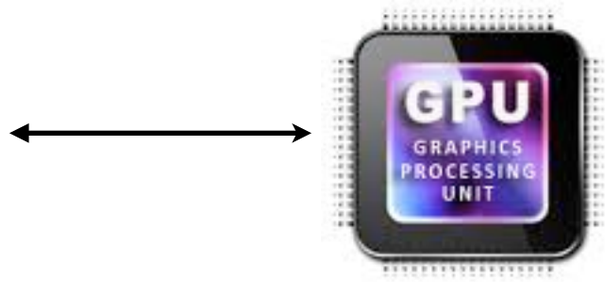
1. to help people to understand & think about data.

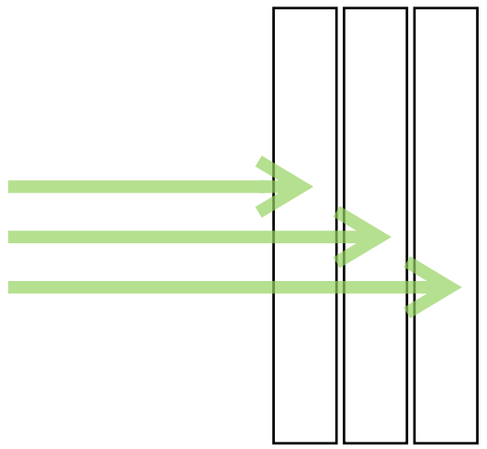
2. to communciate facts



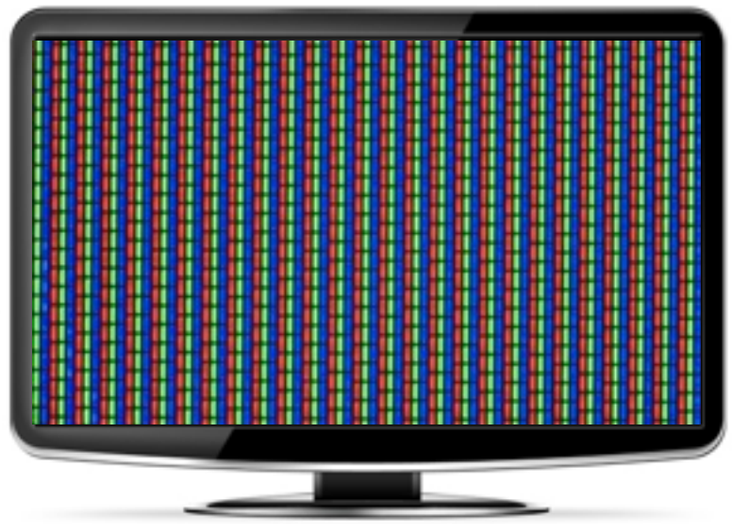
typical computer architecture





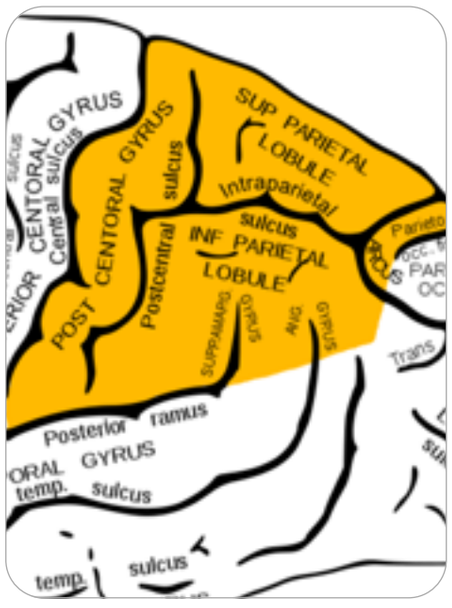


framebuffer(s)



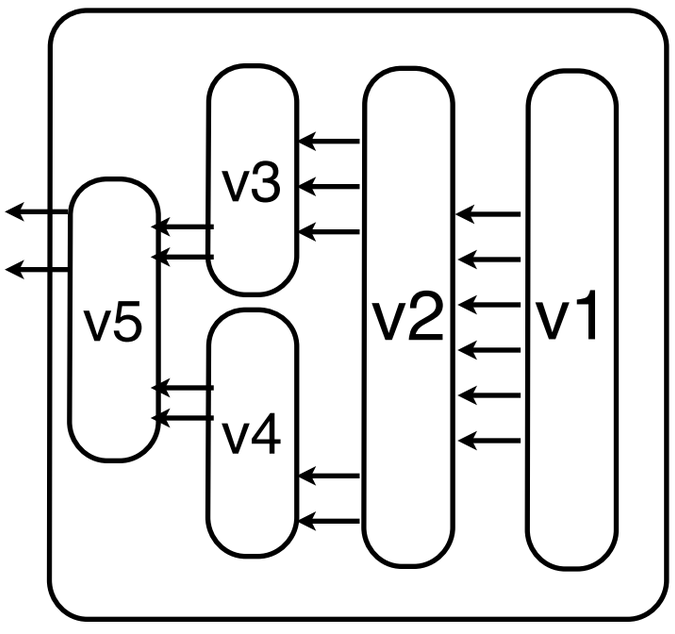
display

parietal lobe + frontal cortex

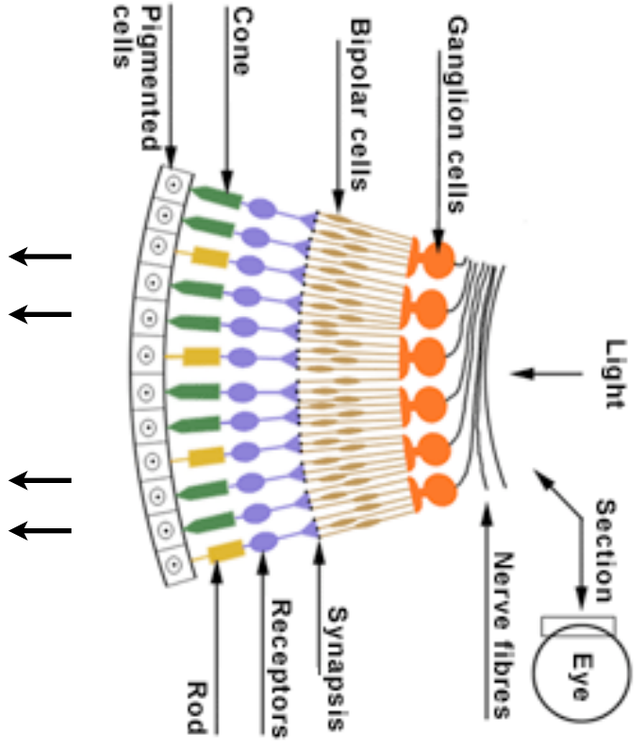


spatial orientation
focus of attention
eye control,
perceptual fusion

occipital lobe

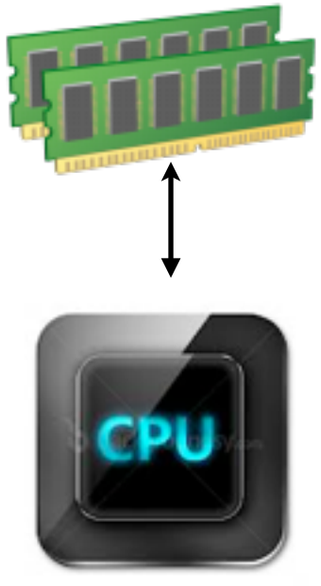


visual cortex
(pattern detection)



retina
(sensing)

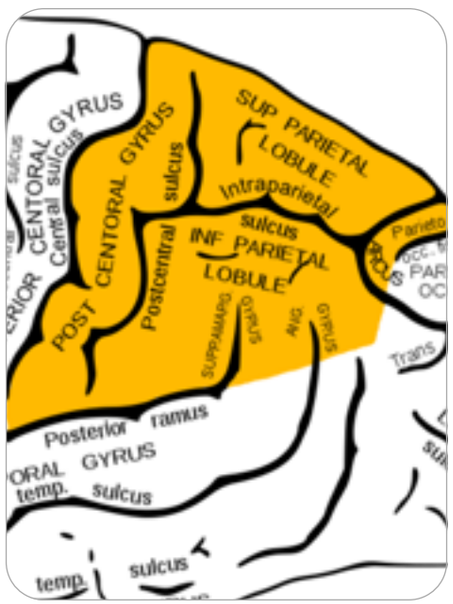
eye / iris / fovea



serial / deliberative processing “attention-focused”

highly parallel

parietal lobe + frontal cortex

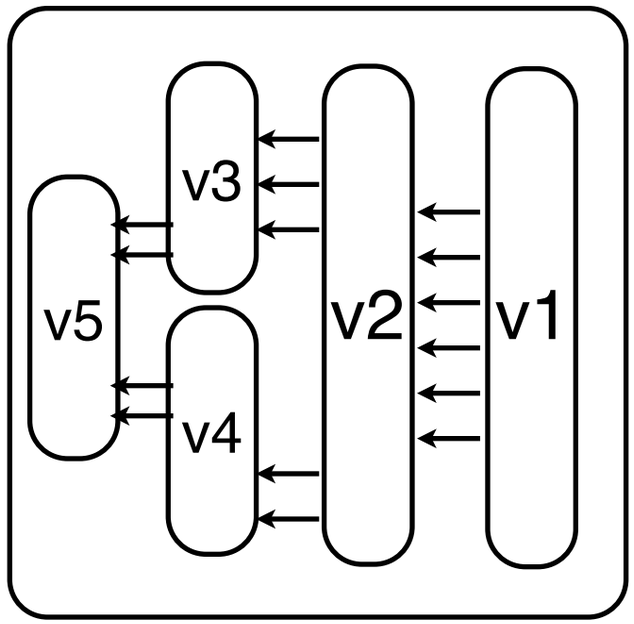


spatial orientation
focus of attention
eye control,
perceptual fusion

access to long term memory

visual processing routines optimised for purpose

occipital lobe



visual cortex
(pattern detection)

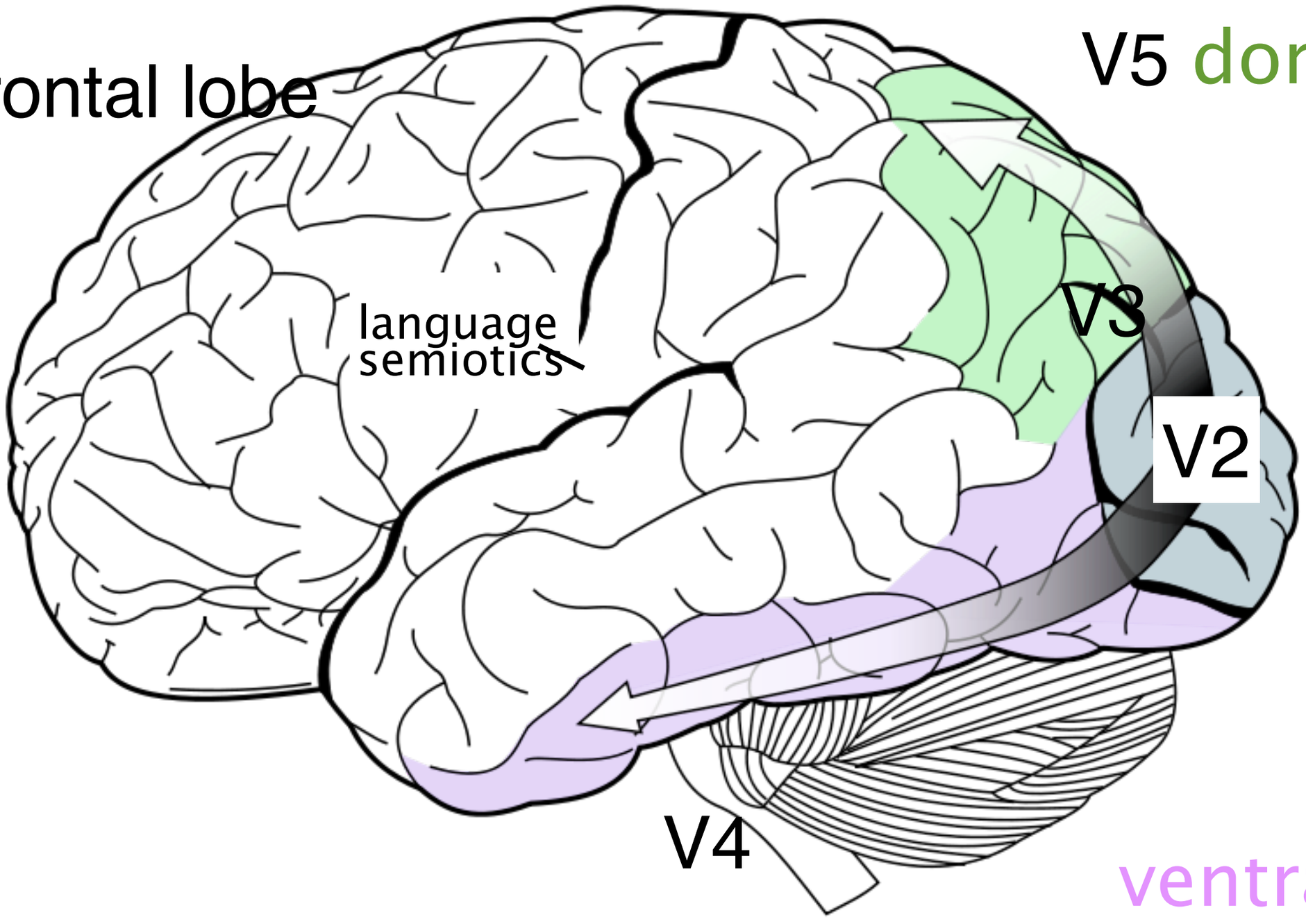
planning
thinking
deliberation
action

spatial reasoning
perceptual fusion
parietal lobe

frontal lobe

V5 dorsal stream
"where/how"
pathway

language
semiotics

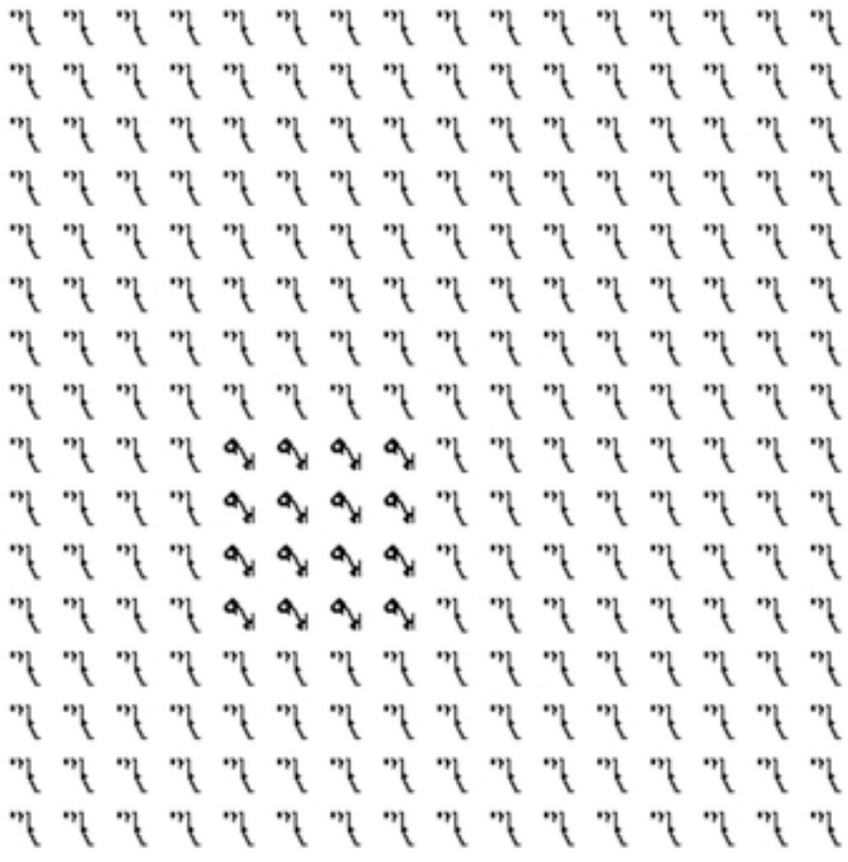


occipital
lobe

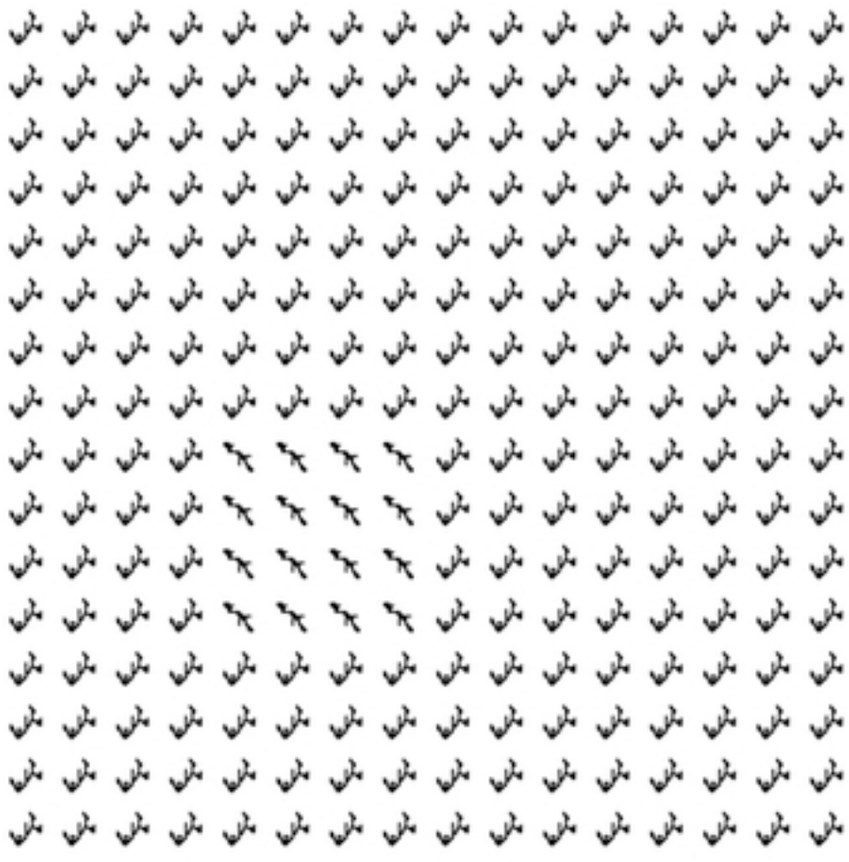
ventral stream
"what" pathway

object and person recognition

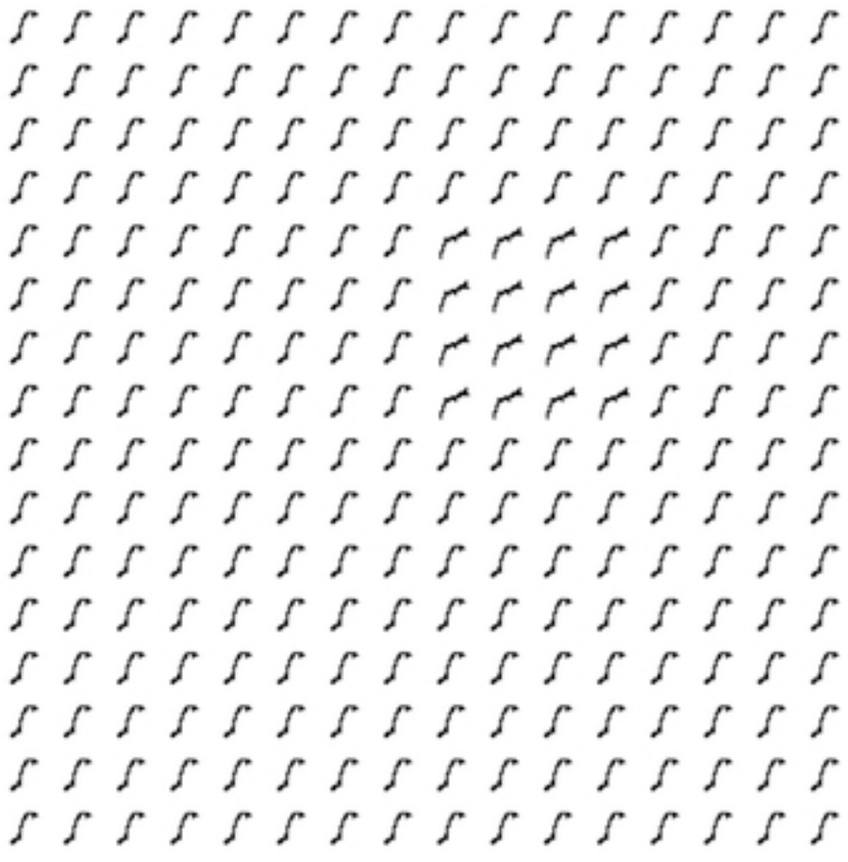
a



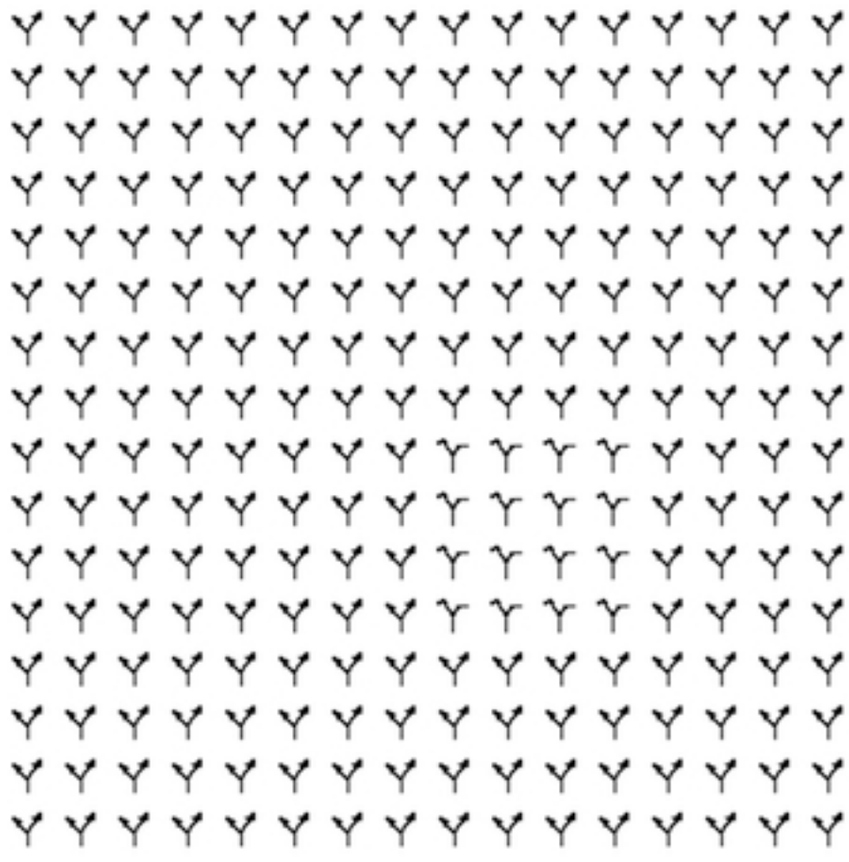
b

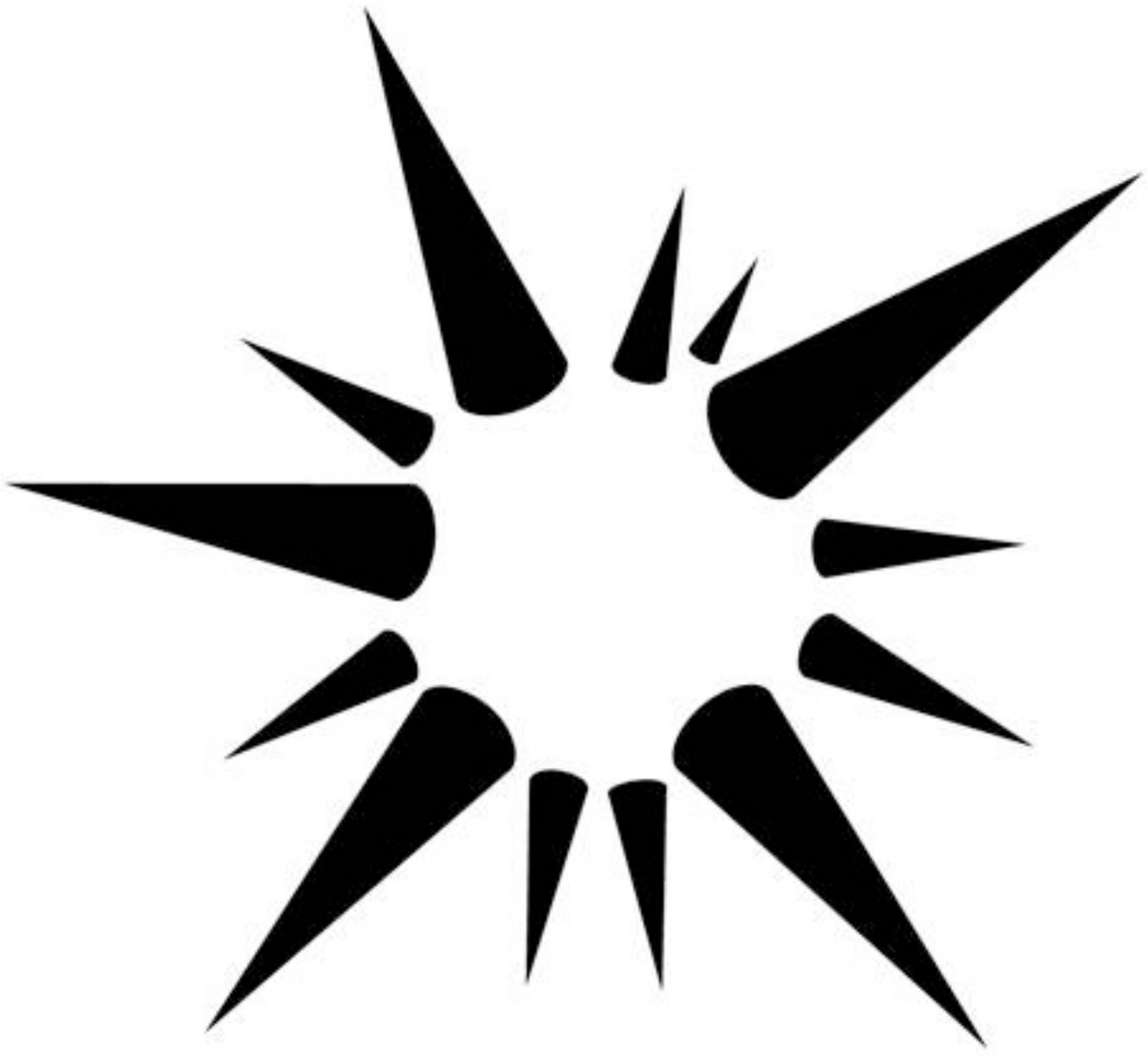


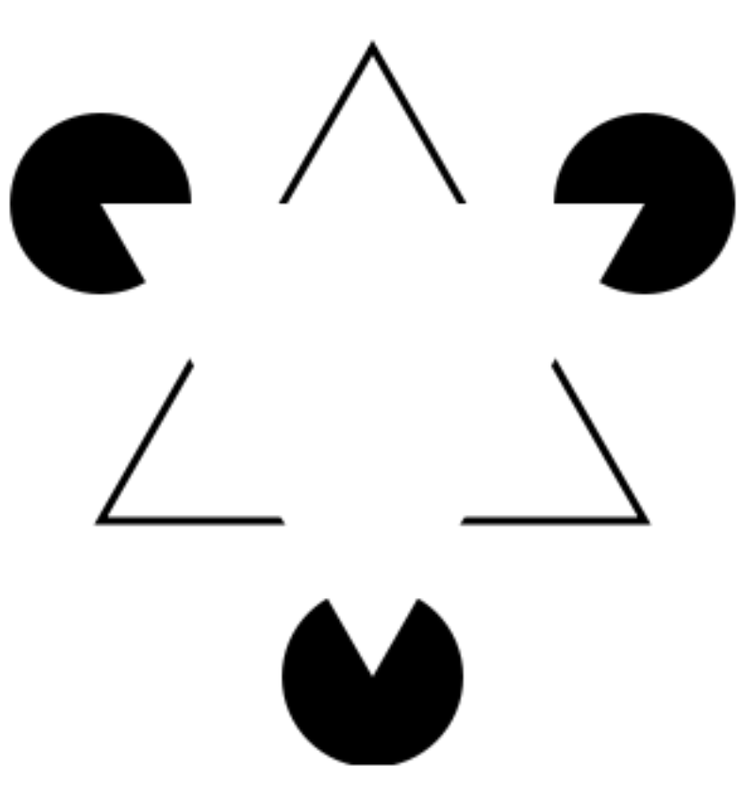
c

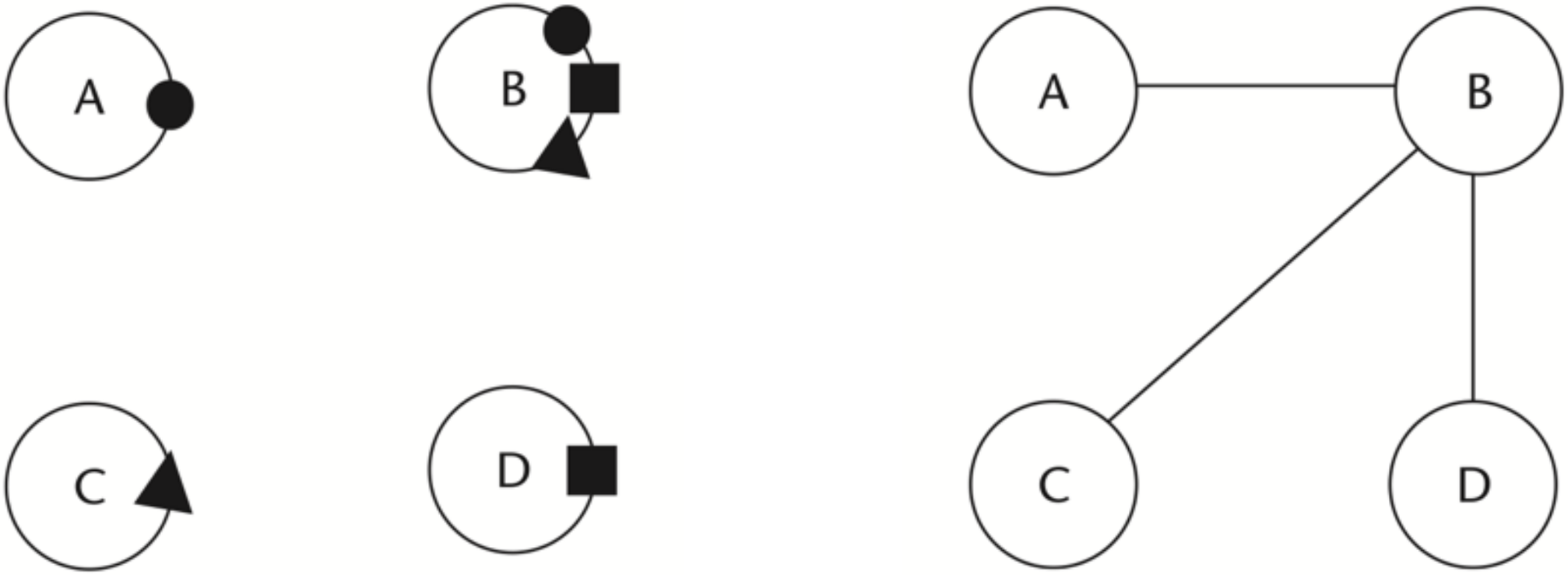


d








A
B
C
D
A
B
C
D

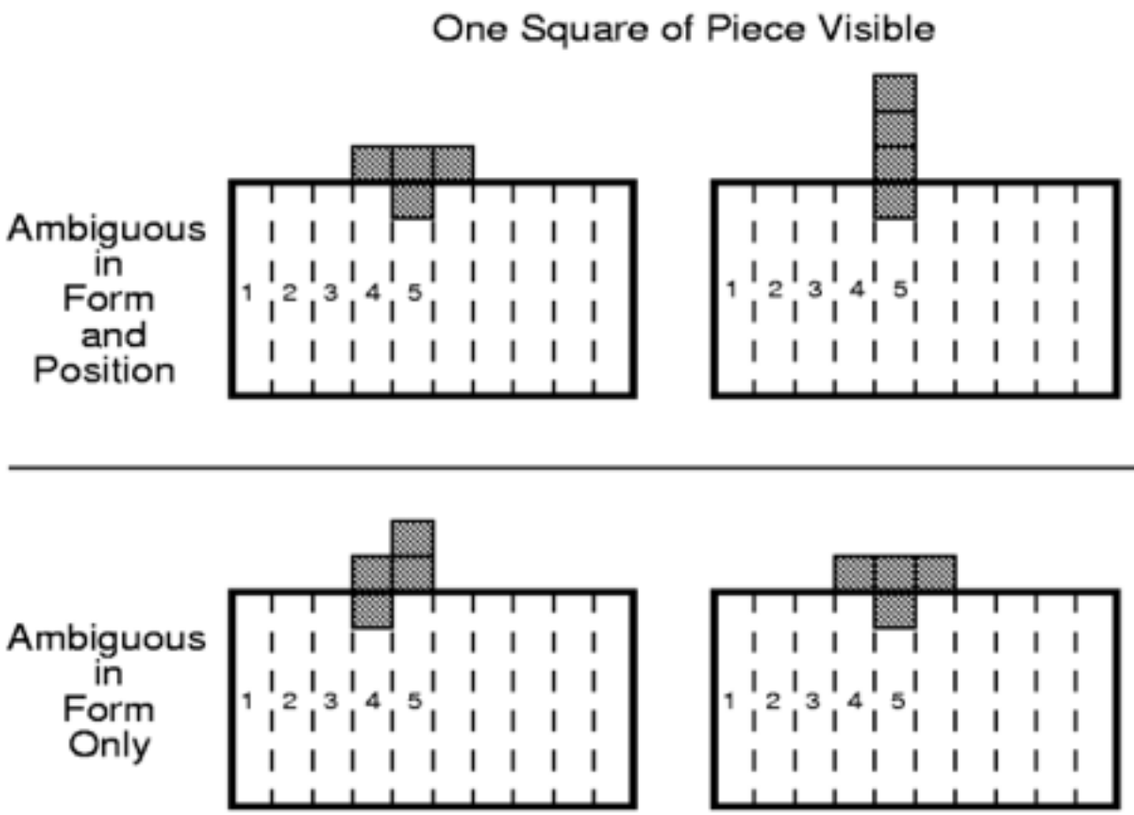


Figure 2a

Figure 2b

London Cholera Outbreak
John Snow, 1854

31 Aug 1854 - 127 deaths in 3
10 Sept - 500 deaths
End of outbreak - 616 deaths

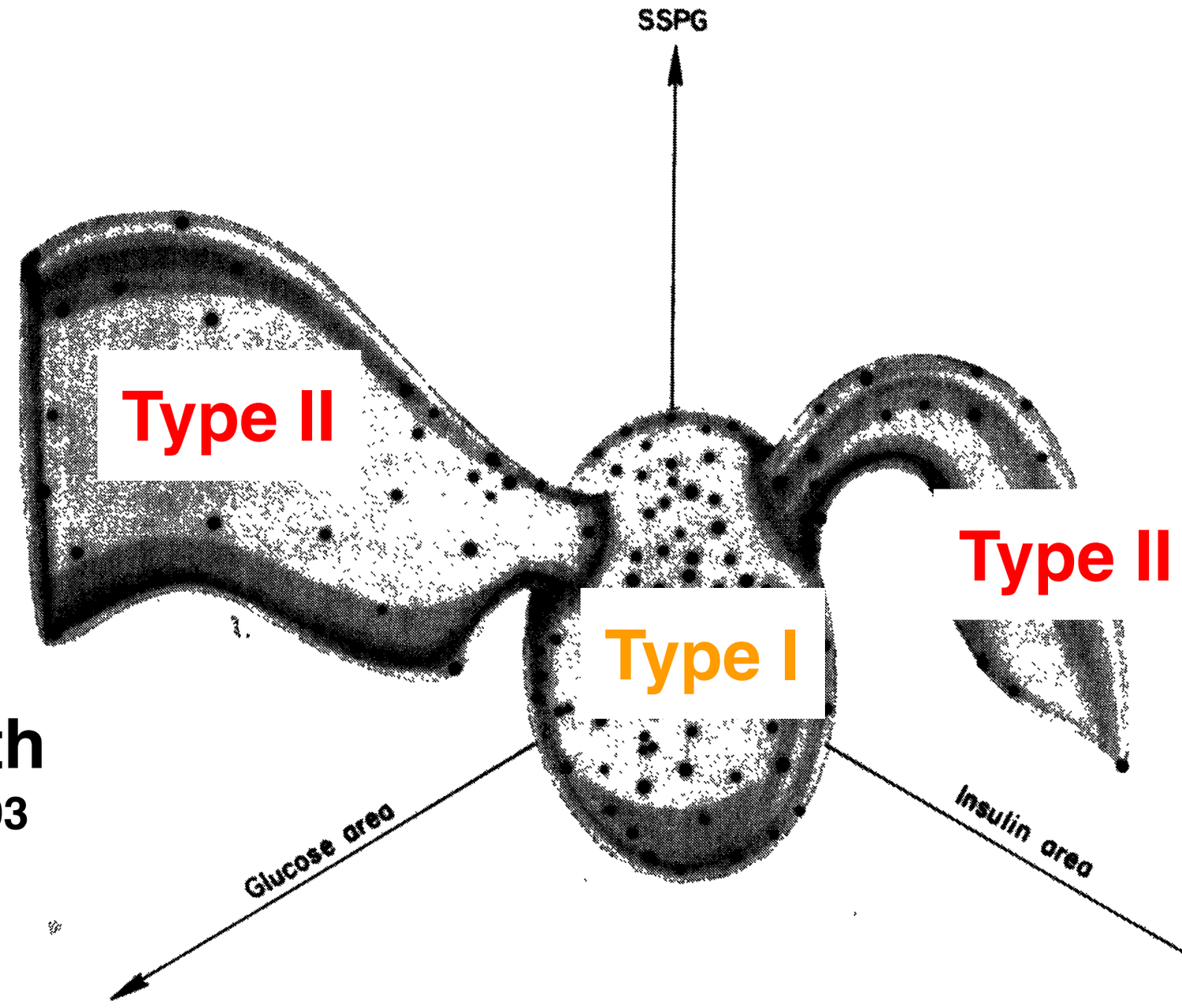
"There was one significant anomaly - none of the monks in the adjacent monastery contracted cholera.

The Story of London's Most Terrifying Epidemic – and How it Changed Science, Cities and the Modern World.

Sir Harold Himsworth

19 May 1905 – 1 November 1993

steady state plasma glucose (response)



glucose area under curve

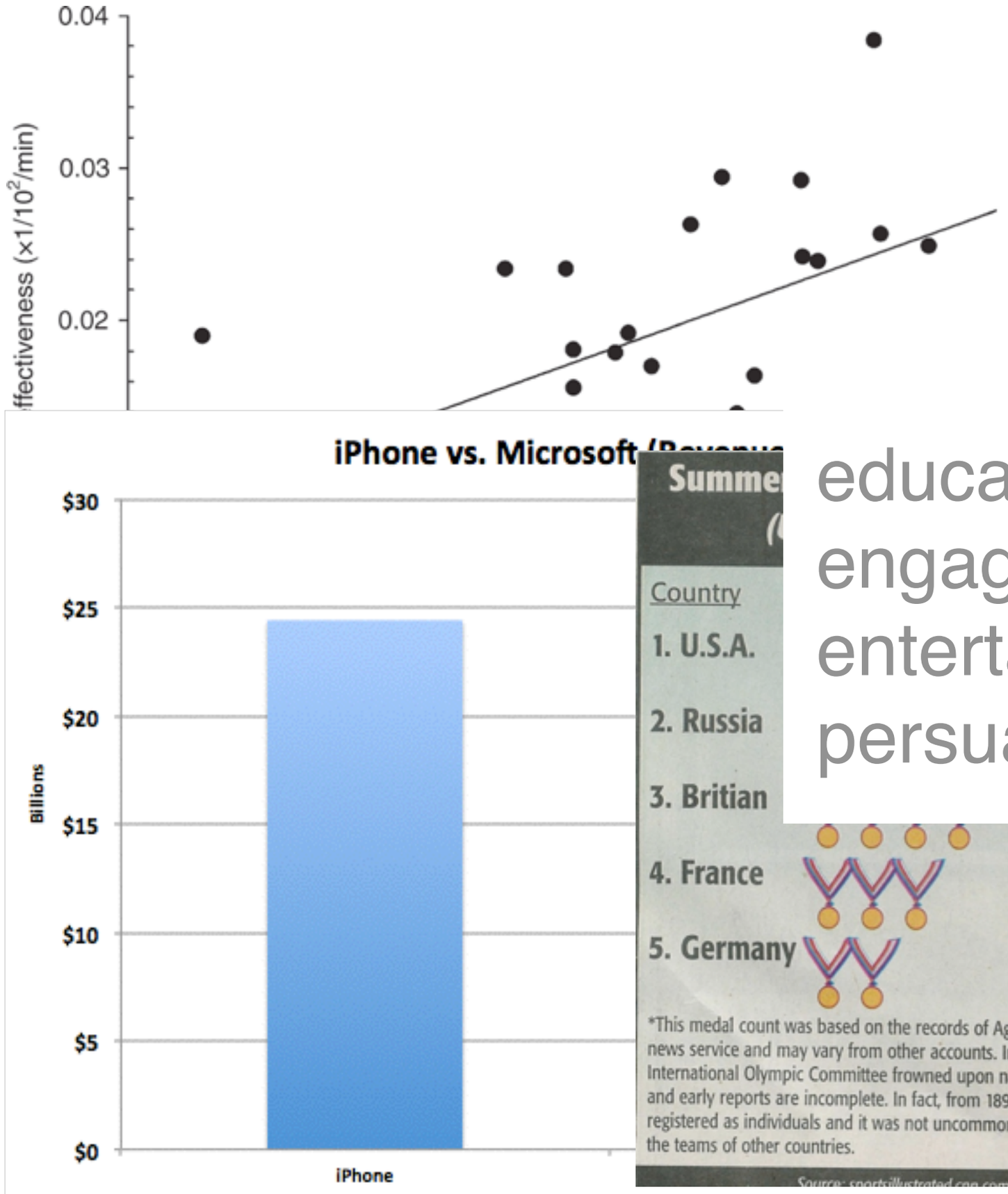
insulin area under curve

so how do we choose appropriate visual representations for our data?

1. purpose

understand

communicate



educate
engage
entertain
persuade

2. data types

{X1, X2, X3, X4, ... }	Xi is...
{1, 200, 5, 6, ... }	integral
{1.0, 2.0, 1.2, 4, ... }	fixed point
{'a', 'b', '12c', 'd' ...}	alpha(-numeric)
{20%, 30%, 1%, 5% ...}	fractional
{🍐,🍎,🥝,🍍,🍍 ...}	categorical
{ f(🍐,🍎),g(🍎,🥝), q(🥝,🍍)...}	relational

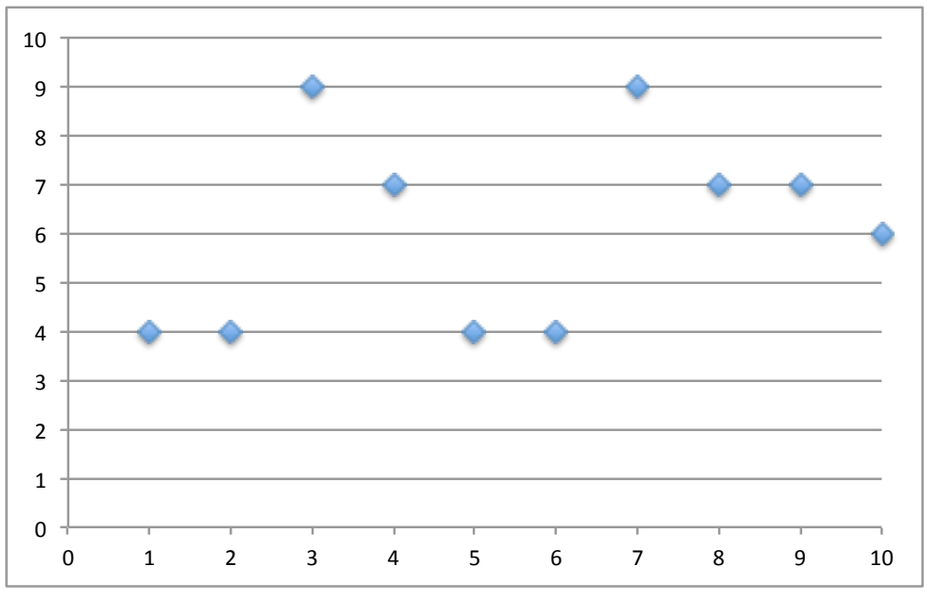
<u>understanding objective</u> - help the user to understand relationships among the elements of the set

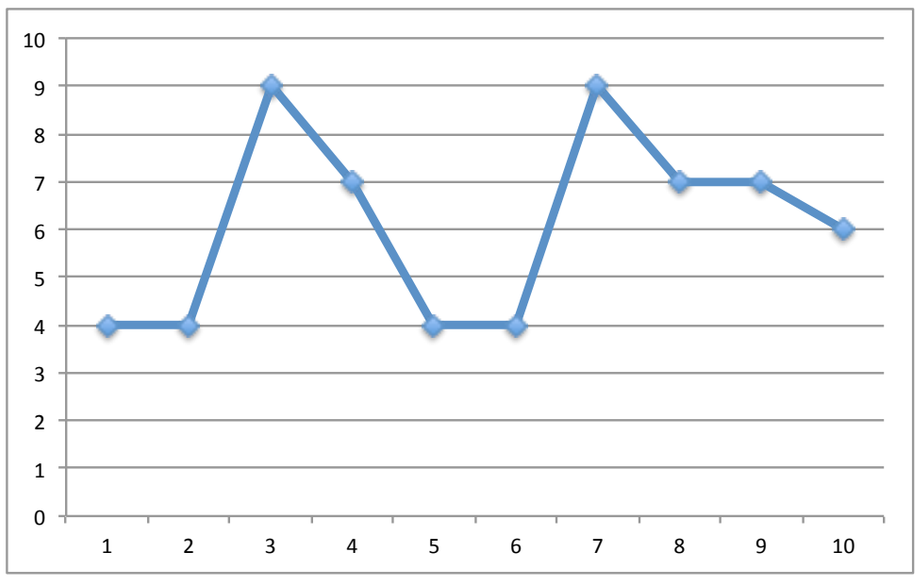
2. data types

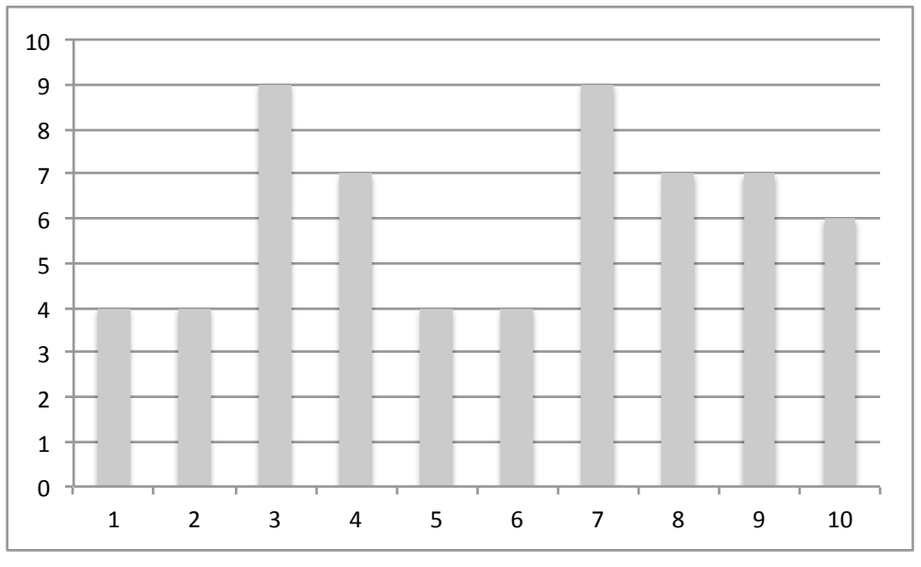
$\{x_1, x_2, x_3, x_4, \ldots\}$	x_i is...
$\{1, 200, 5, 6, \ldots\}$	integral
$\{1.0, 2.0, 1.2, 4, \ldots\}$	fixed point
$\{$'a', 'b', '12c', 'd' ...$\}$	alpha(-numeric)
$\{20\%, 30\%, 1\%, 5\% \ldots\}$	fractional
{ , , , , ...}	categorical
{ f(,),g(,), q(,) ...}	relational

understanding objective - help the user to understand relationships among the elements of the set

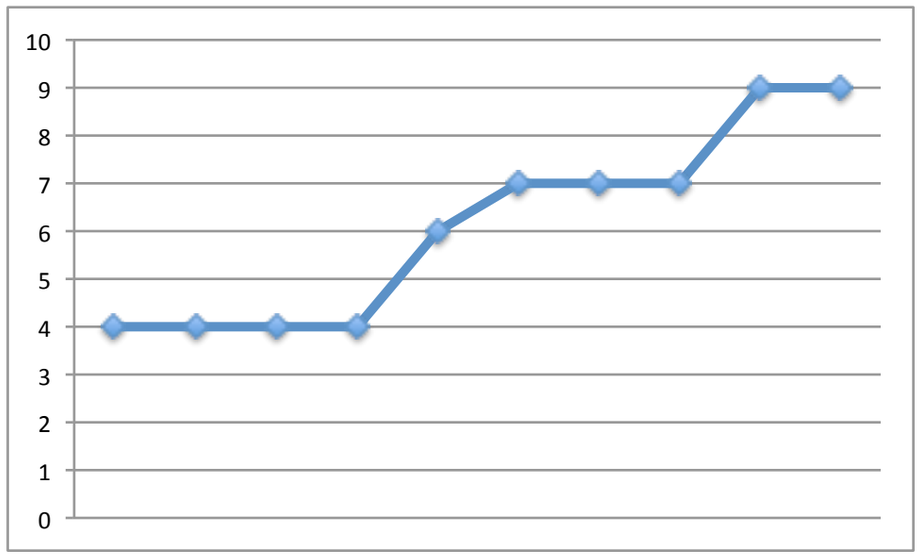
4
4
9
7
4
4
9
7
7
6



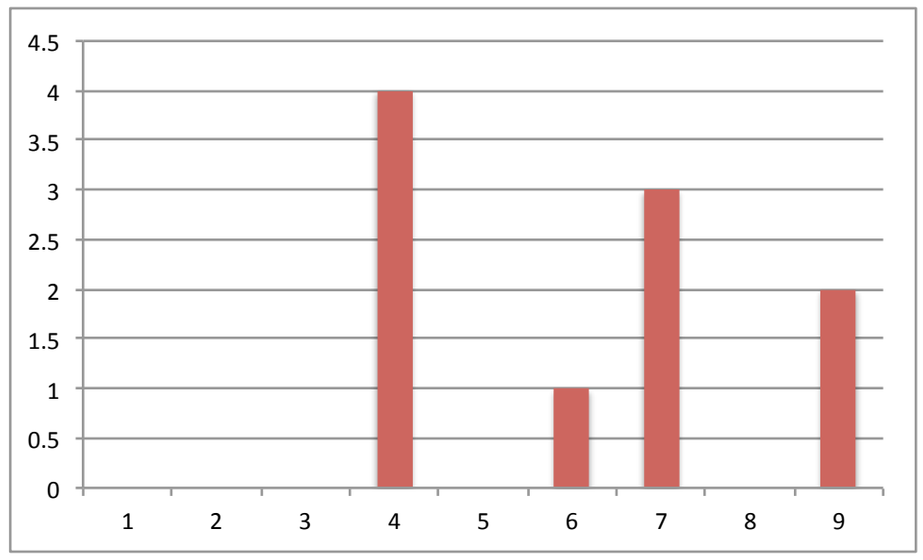




ordering significant



sorted



histogram

box & whisker

order insignificant

so you have a dataset...

it's probably multivariate

$$X = \{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \dots\}$$

$$X = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ \text{🍐}, & \text{🥝}, & \text{🍍} \\ t_1 & t_2 & t_3 \end{bmatrix} \dots$$

if these are observations of the [same] of object(s) over time
"time series"

if these are observations of different things at a single point in time
"population"

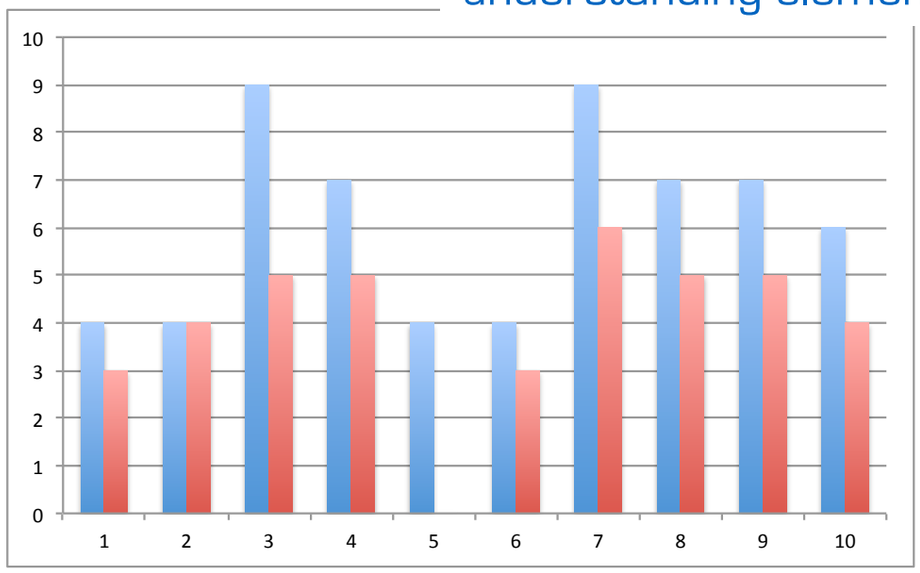
if these are observations of different things at a different points in time
"observations"

<u>understanding objective(s)</u> :

1. relations among dimensions of each sample (multivariate)
2. relations among samples/observations (multidimensional)

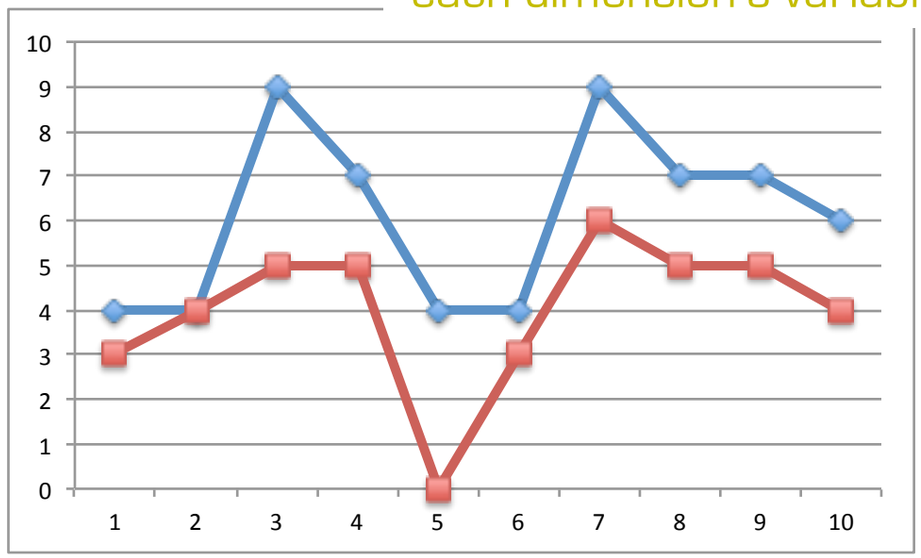
4	3
4	4
9	5
7	5
4	0
4	3
9	6
7	5
7	5
6	4

understanding elements



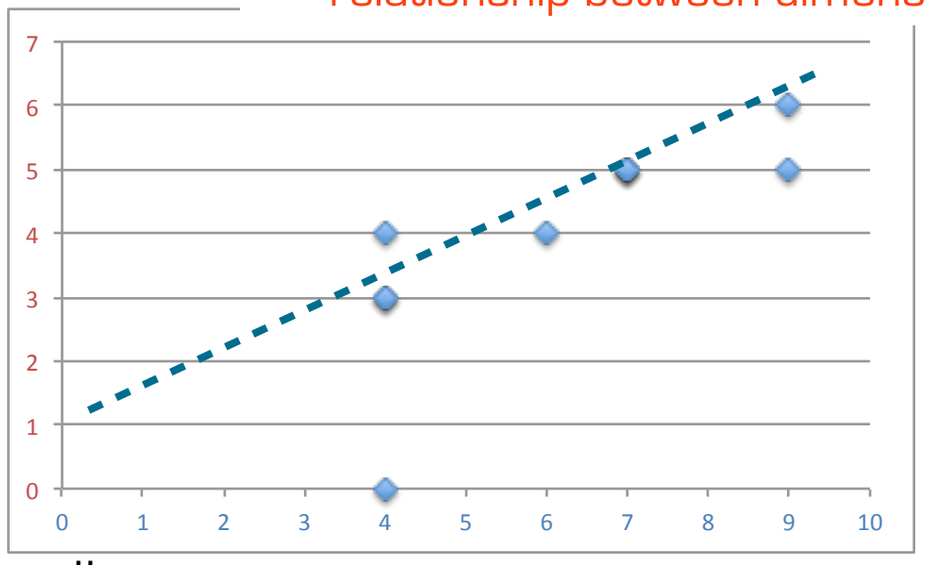
clustered bar

each dimension's variability



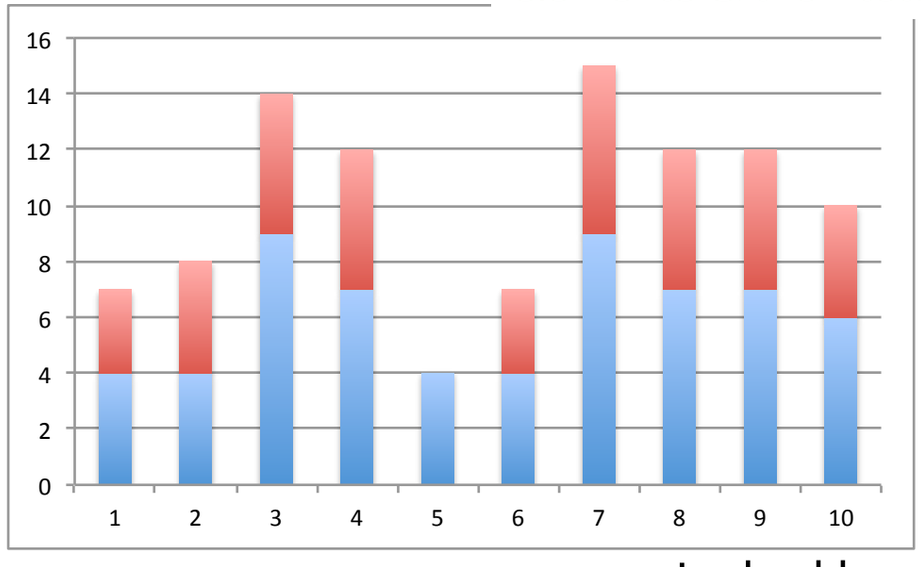
lines

relationship between dimensions



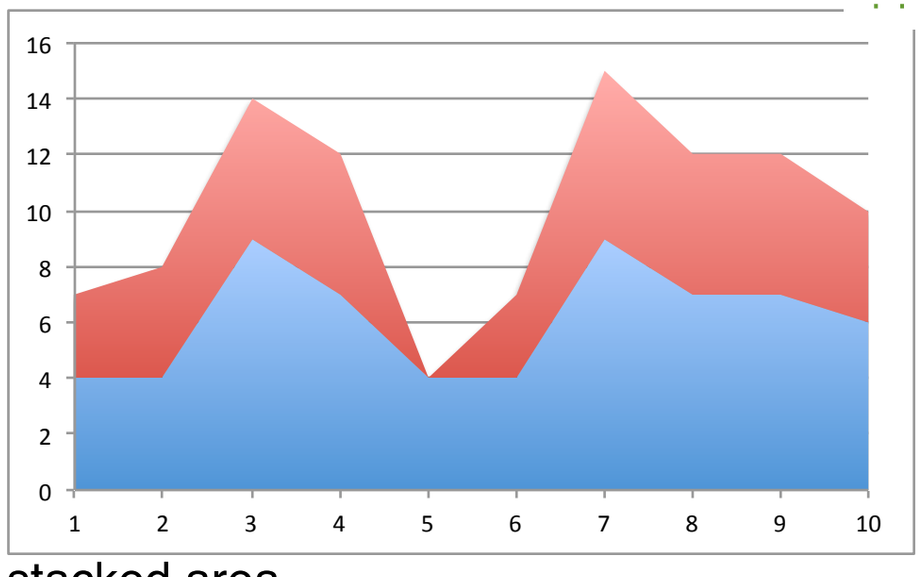
scatter

elements & their totals

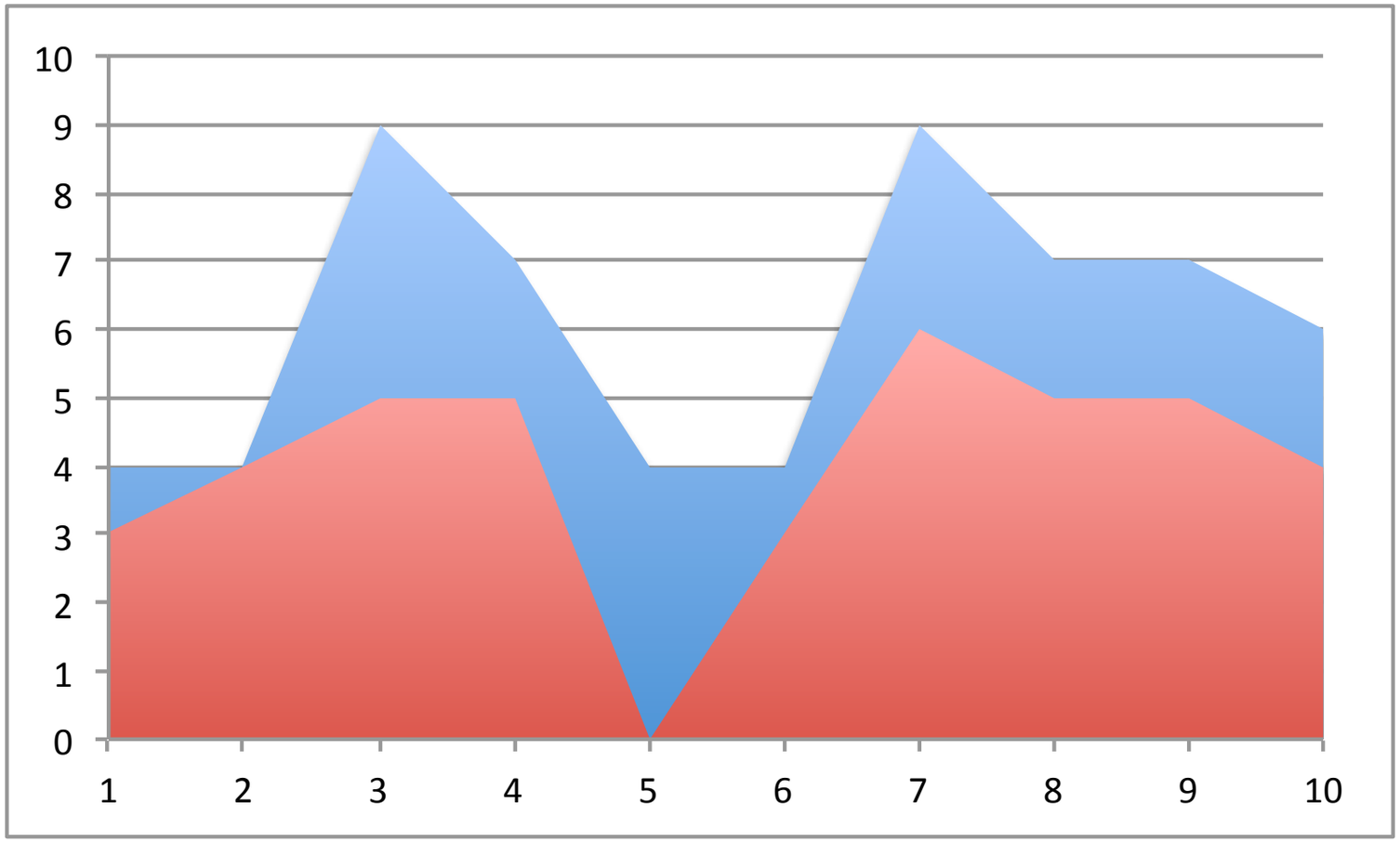


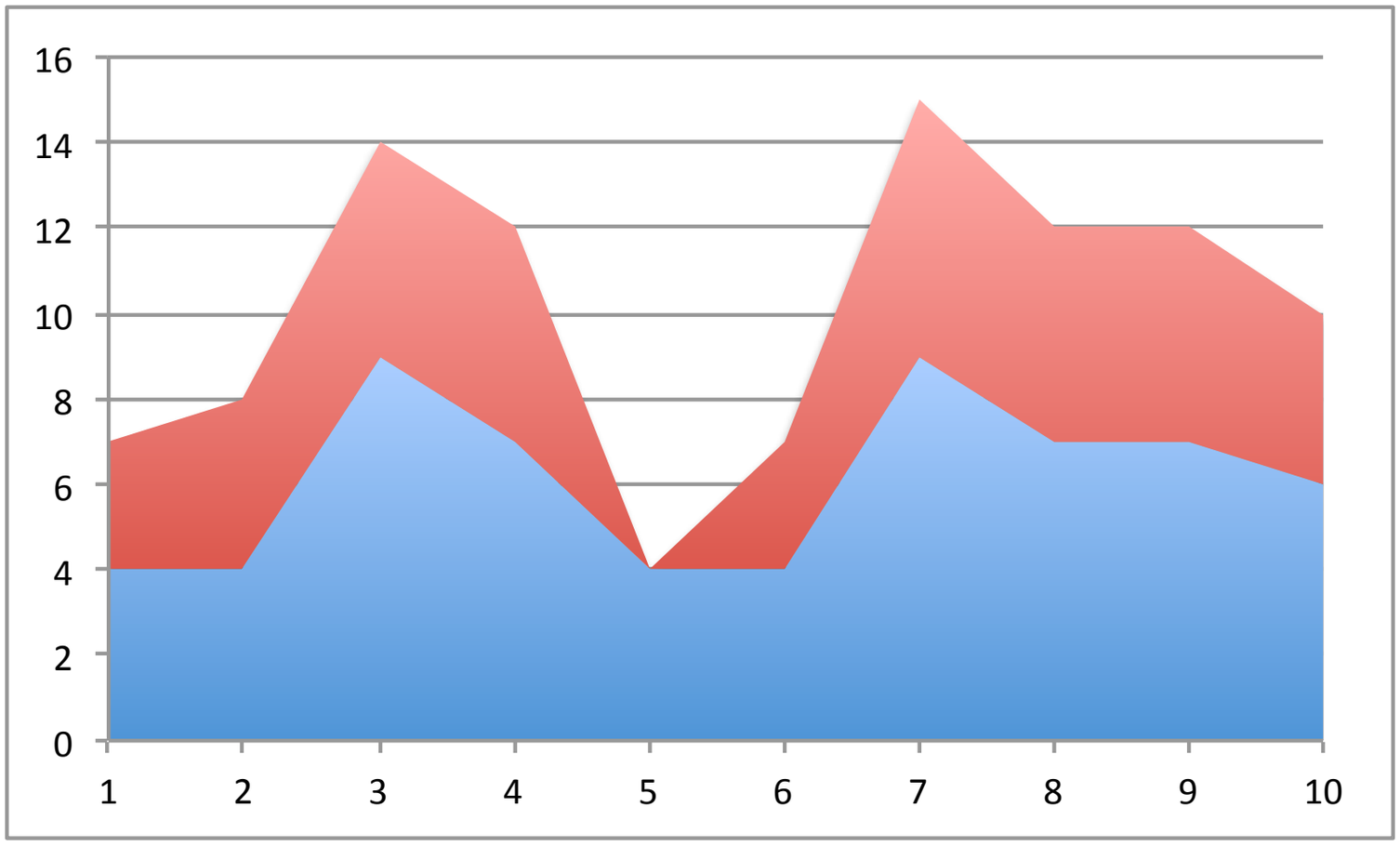
stacked bar

???



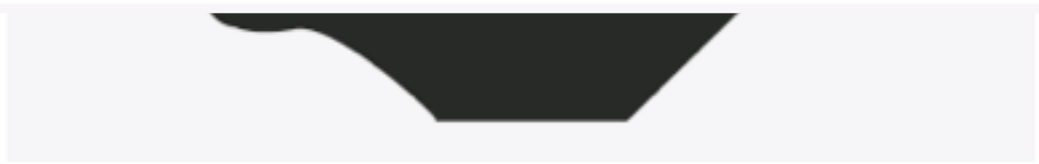
stacked area

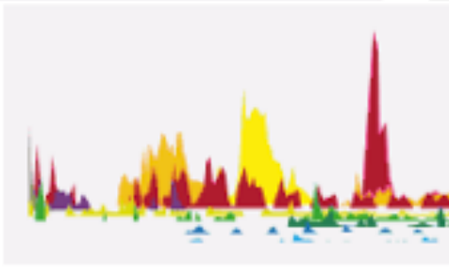

10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10

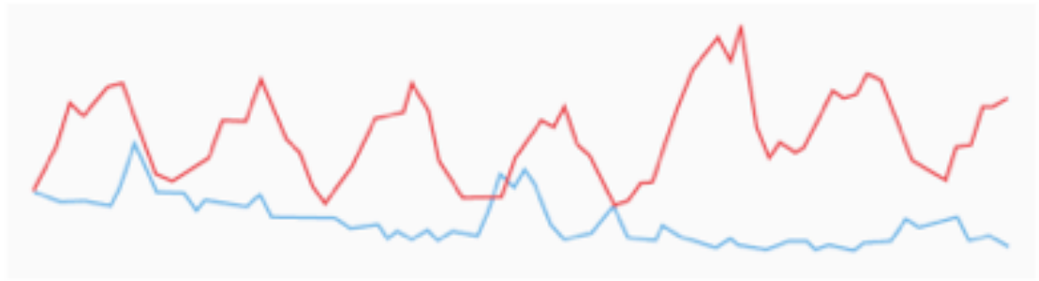

16
14
12
10
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10

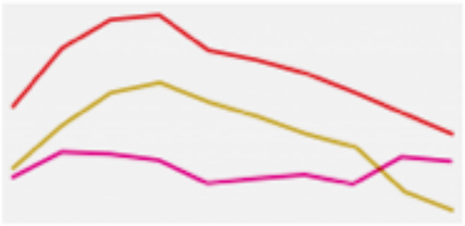
Exercise

- Taking the UK Trade Exports data:
  - How many dimensions?
  - Which dimensions would you pick to visualise for understanding?
  - Which dimensions would you pick to visualise for communication?
  - How do you get to the data for each?









3. Visual Dimensions

data dimension types

integral

fixed point

alpha(-numeric)

ordinals

categorical

relational

...

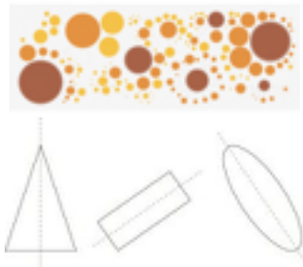
visual dimension type

position — relative location, centrality

shape

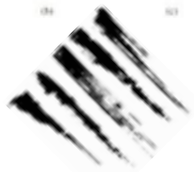
colour — saturation, opacity

size — width, height



orientation

colour, pattern, thickness



stroke

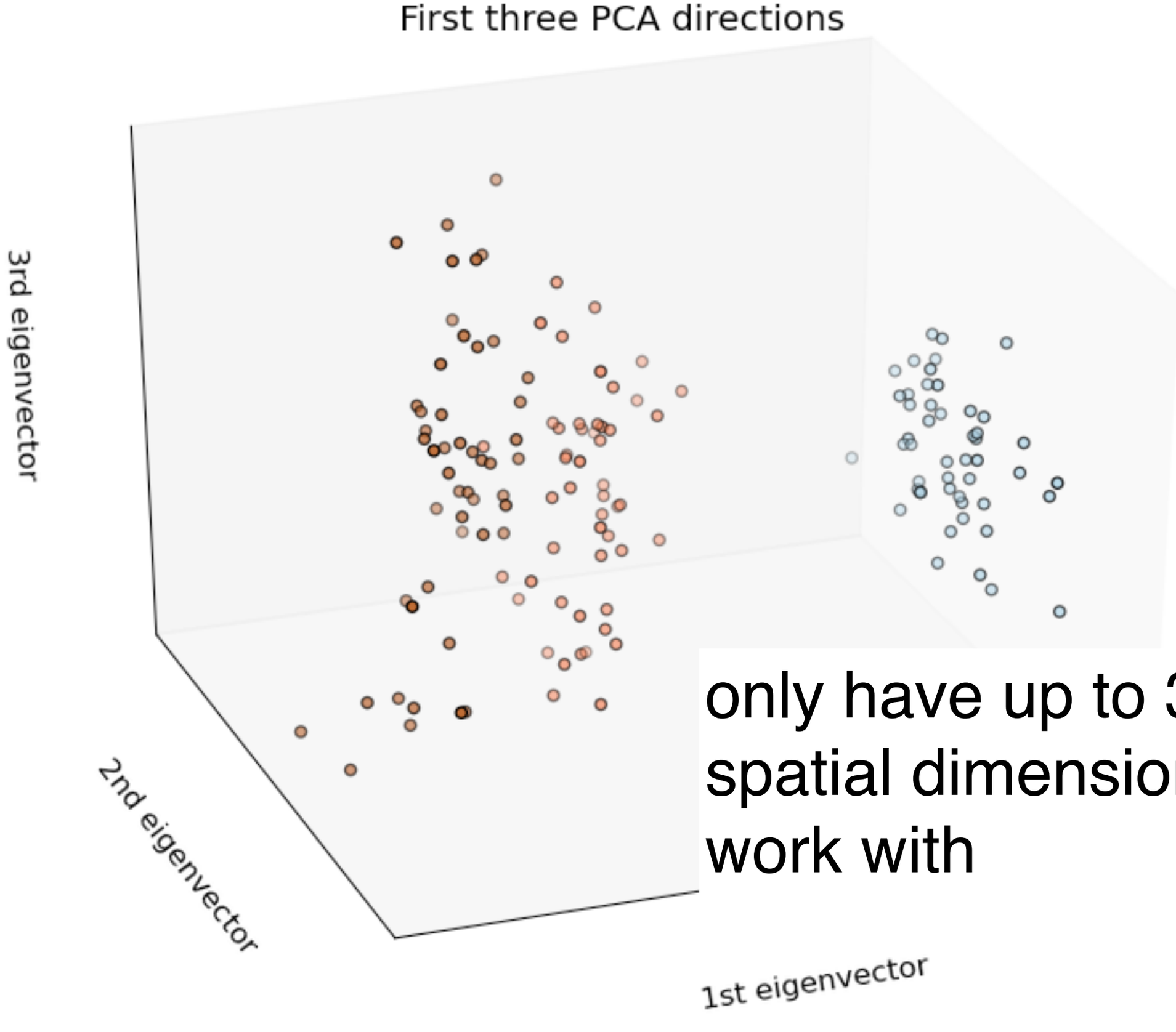
opacity

texture

movement

juxtaposition

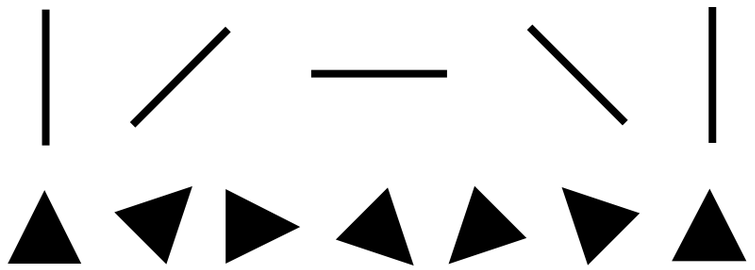
position



only have up to 3 spatial dimensions to work with

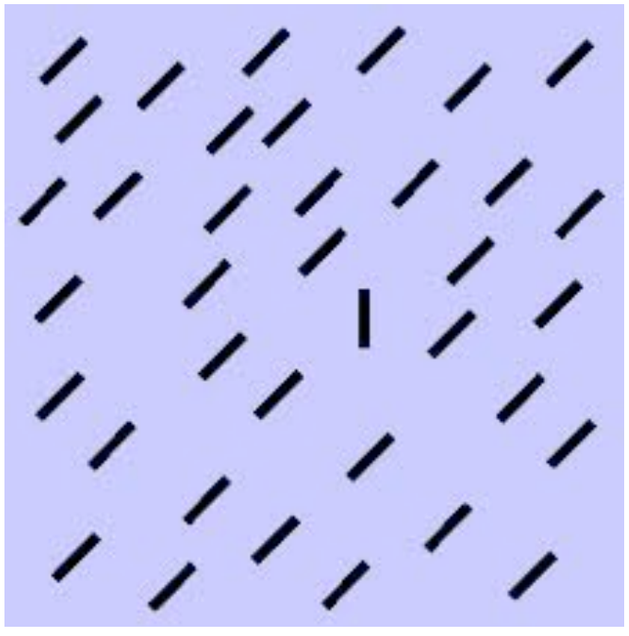
orientation

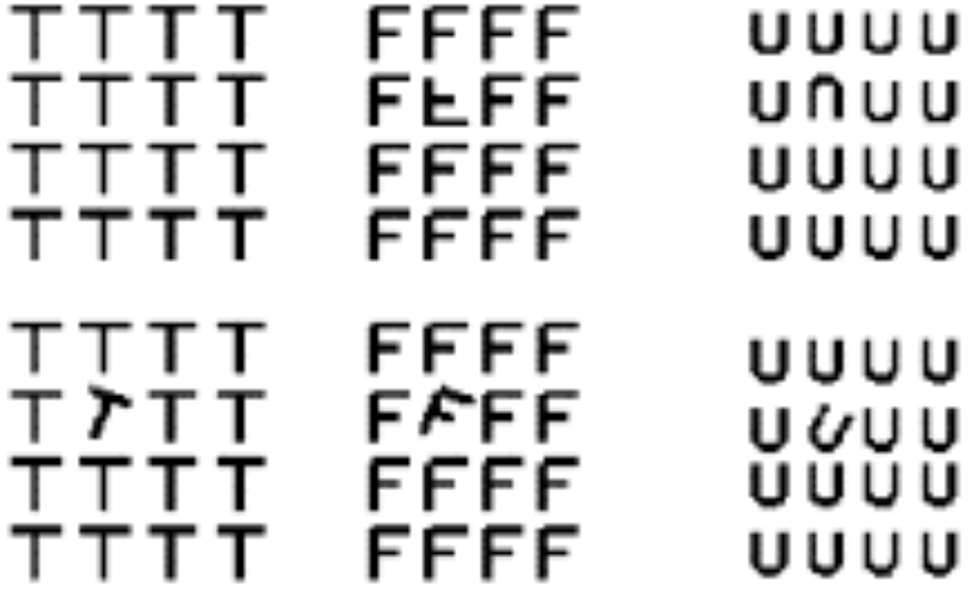
range-limited



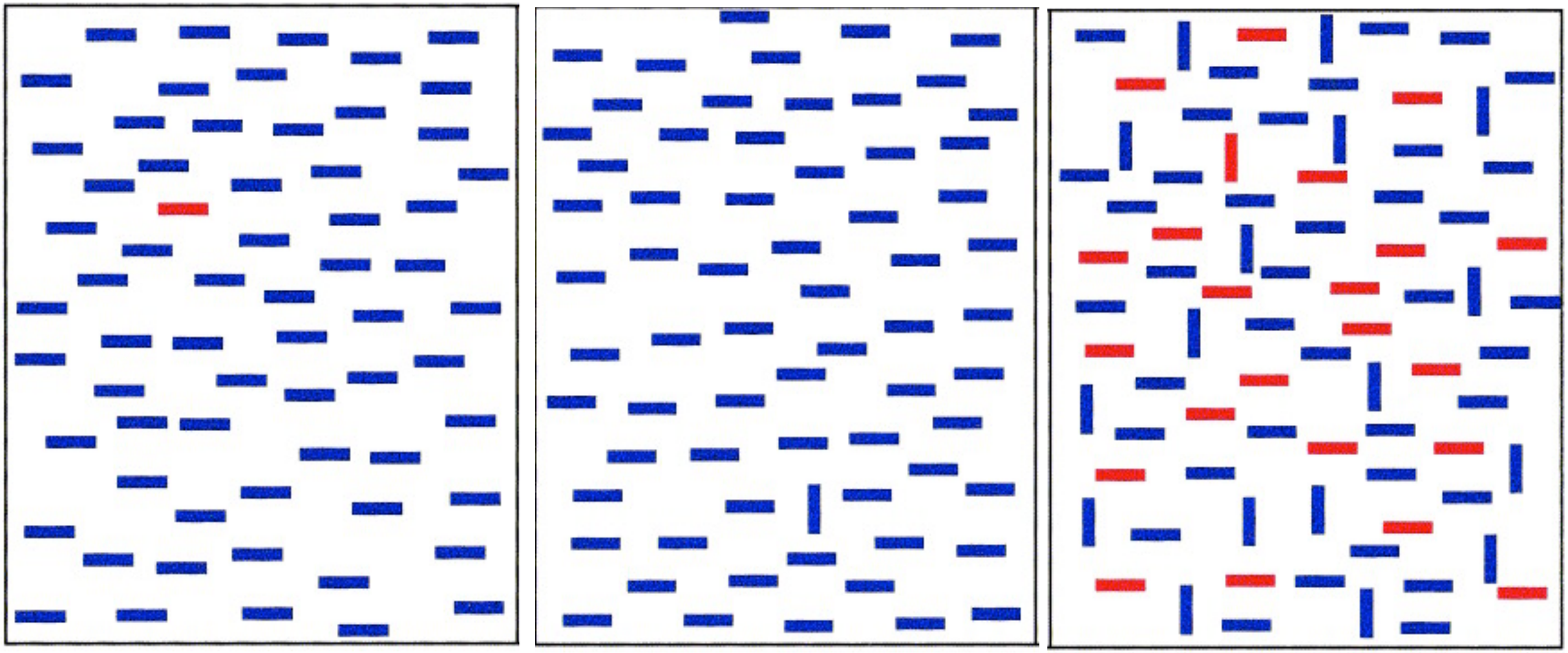
symmetry properties of the geometry

pop-out





orientation
popouts using multiple dimensions

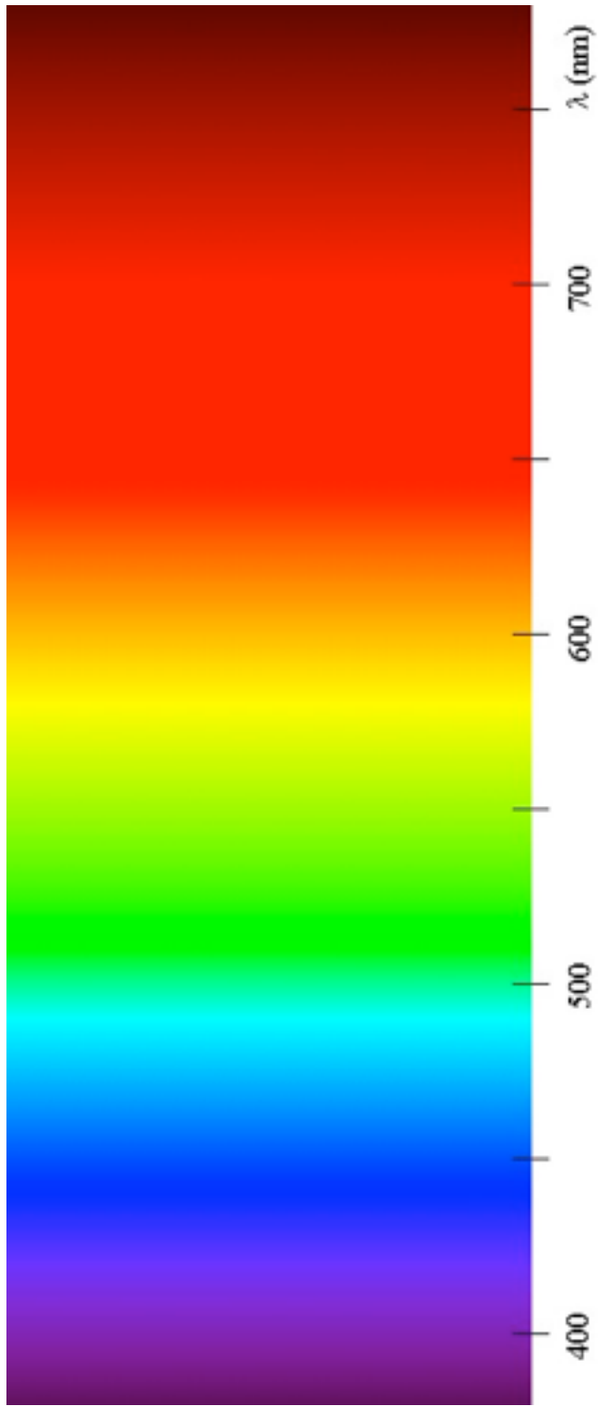


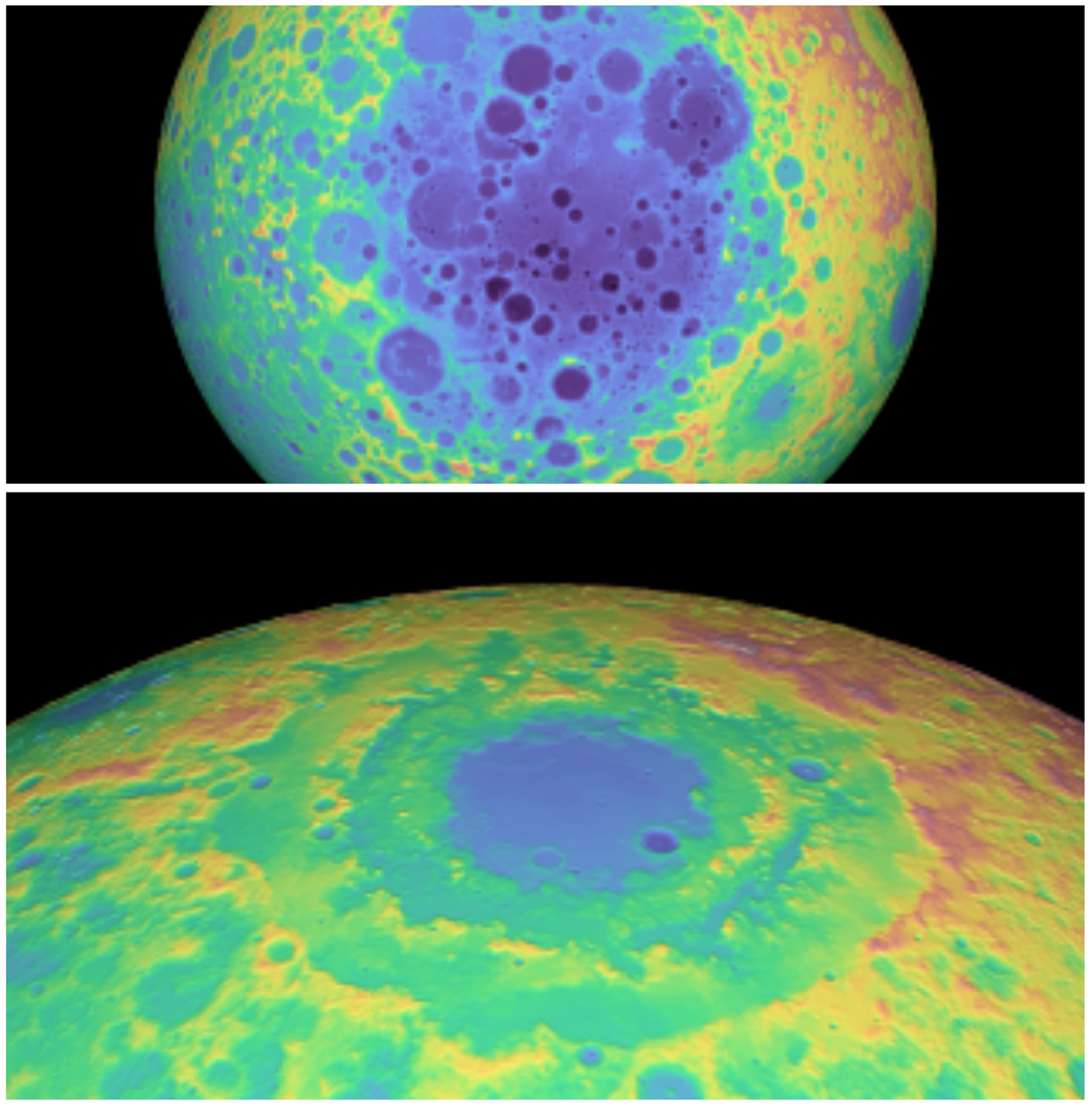
1D colour

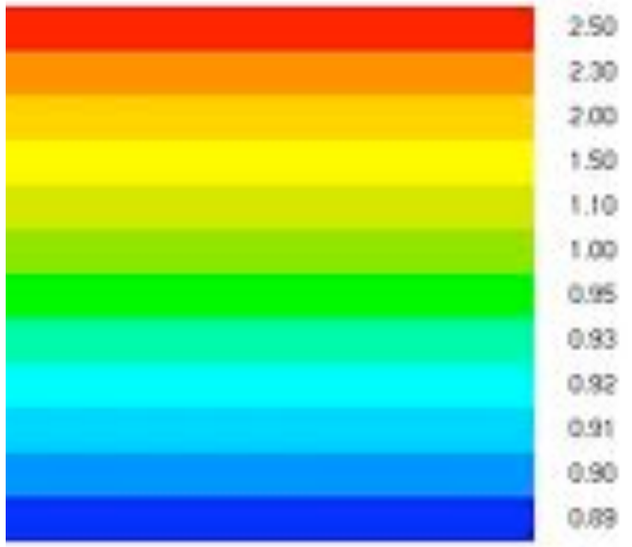
1D orientation

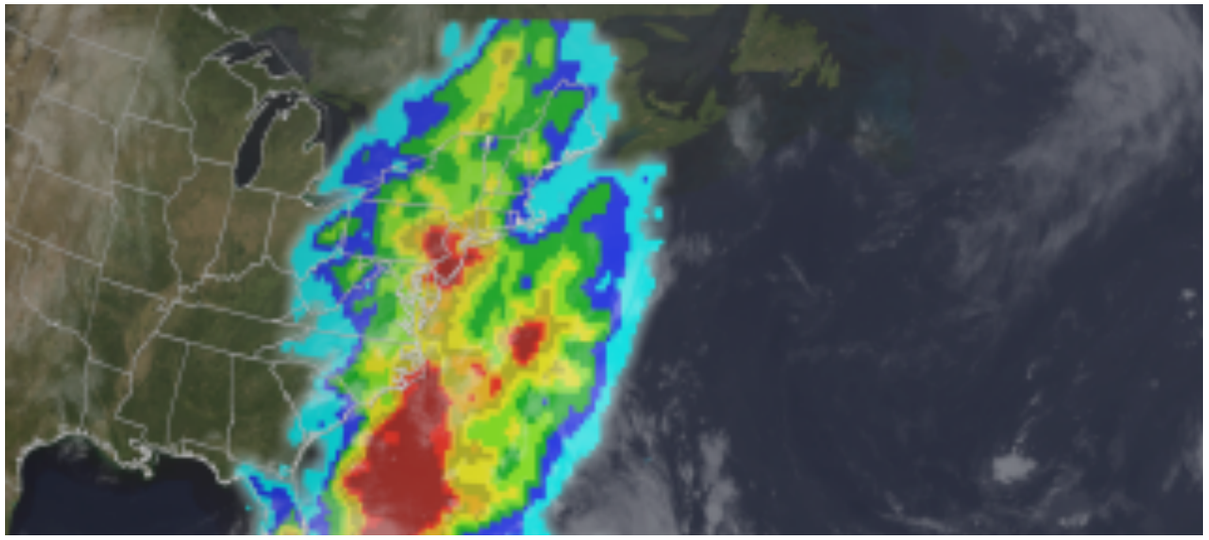
2D color/
orientation

Using colour for continuous values

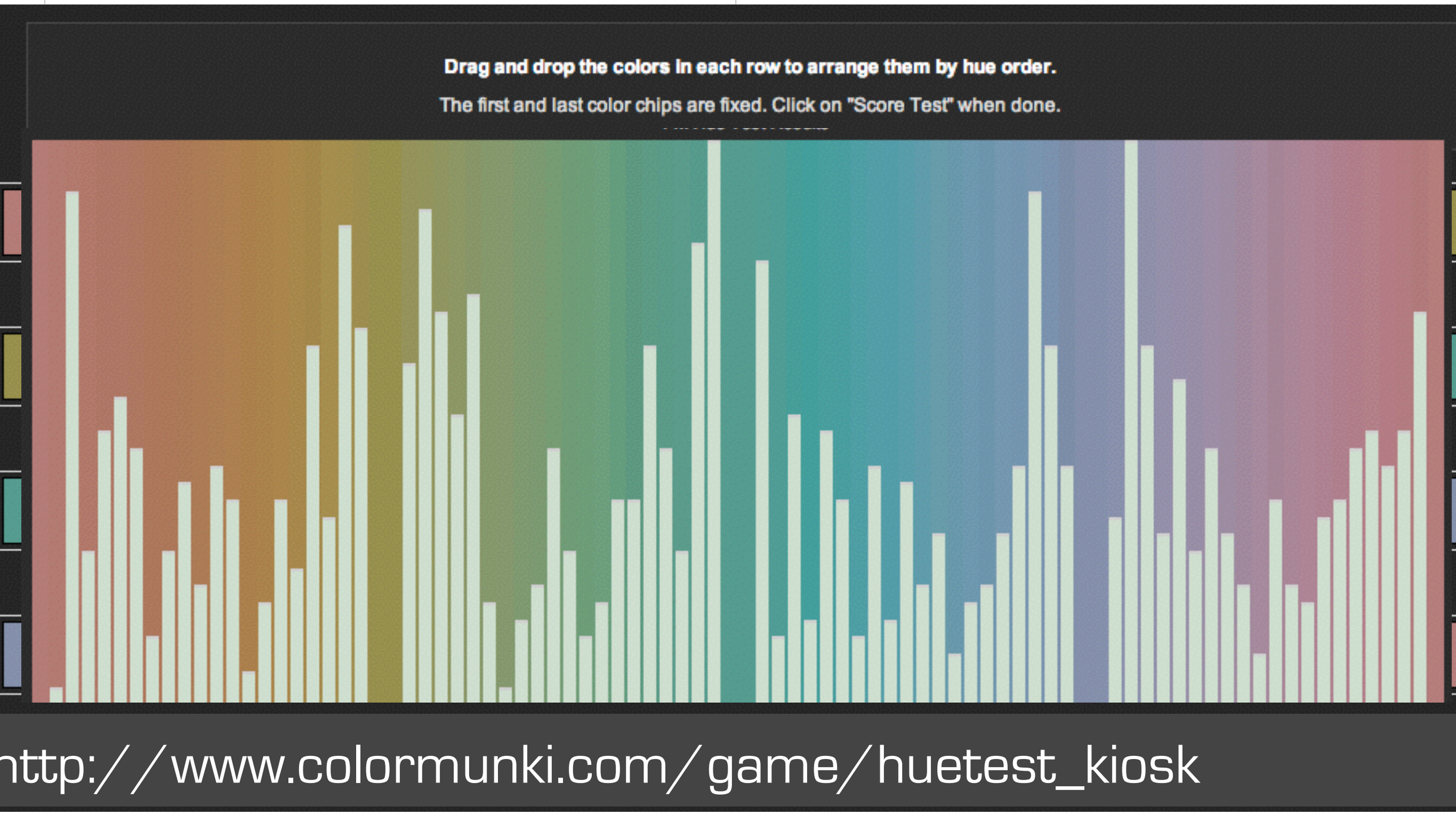






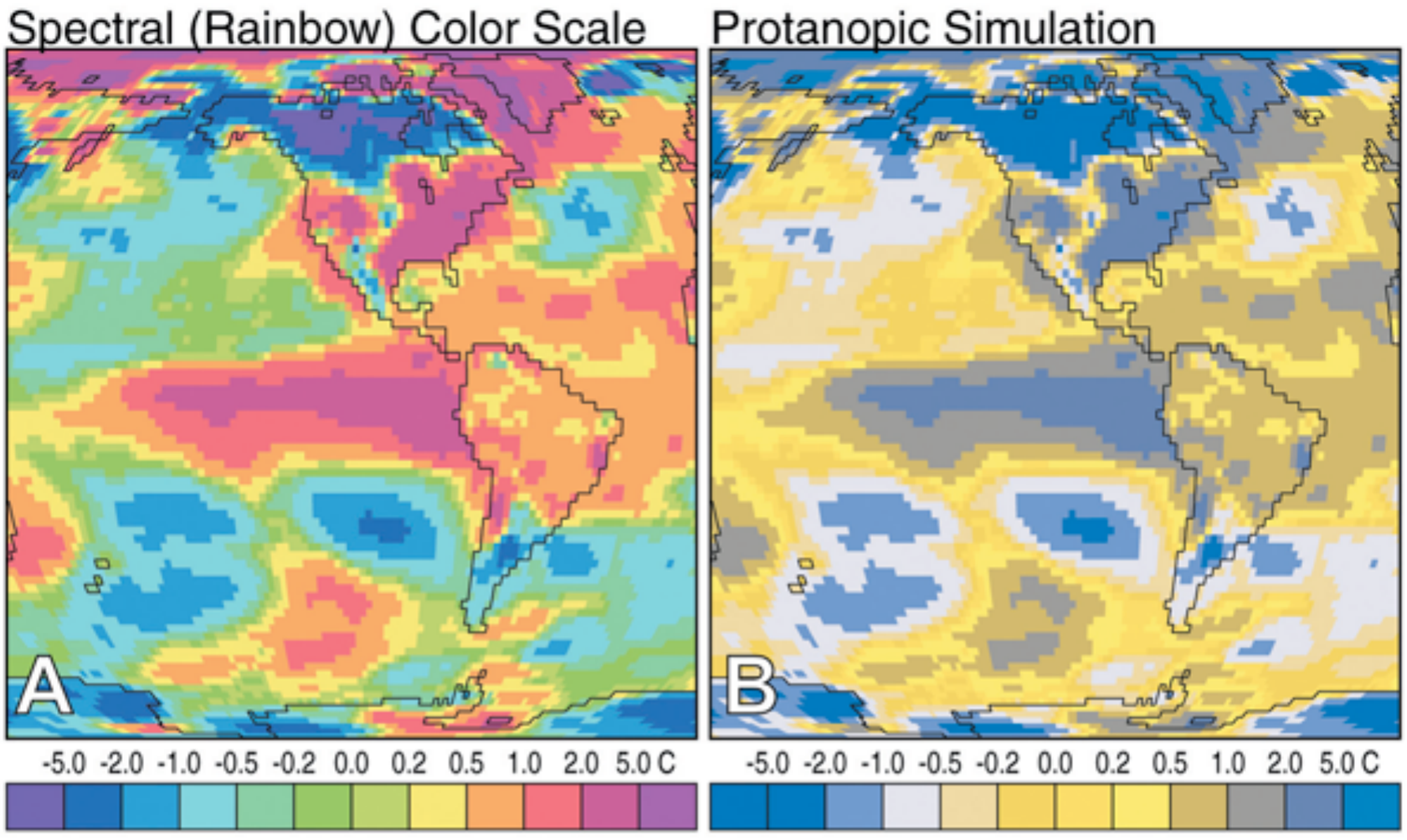


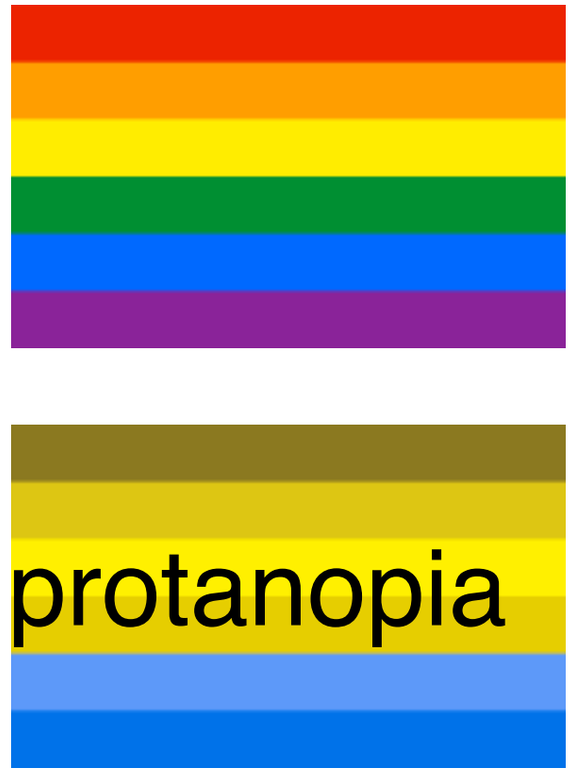
Using colour for continuous values



problem 1: No natural ordering

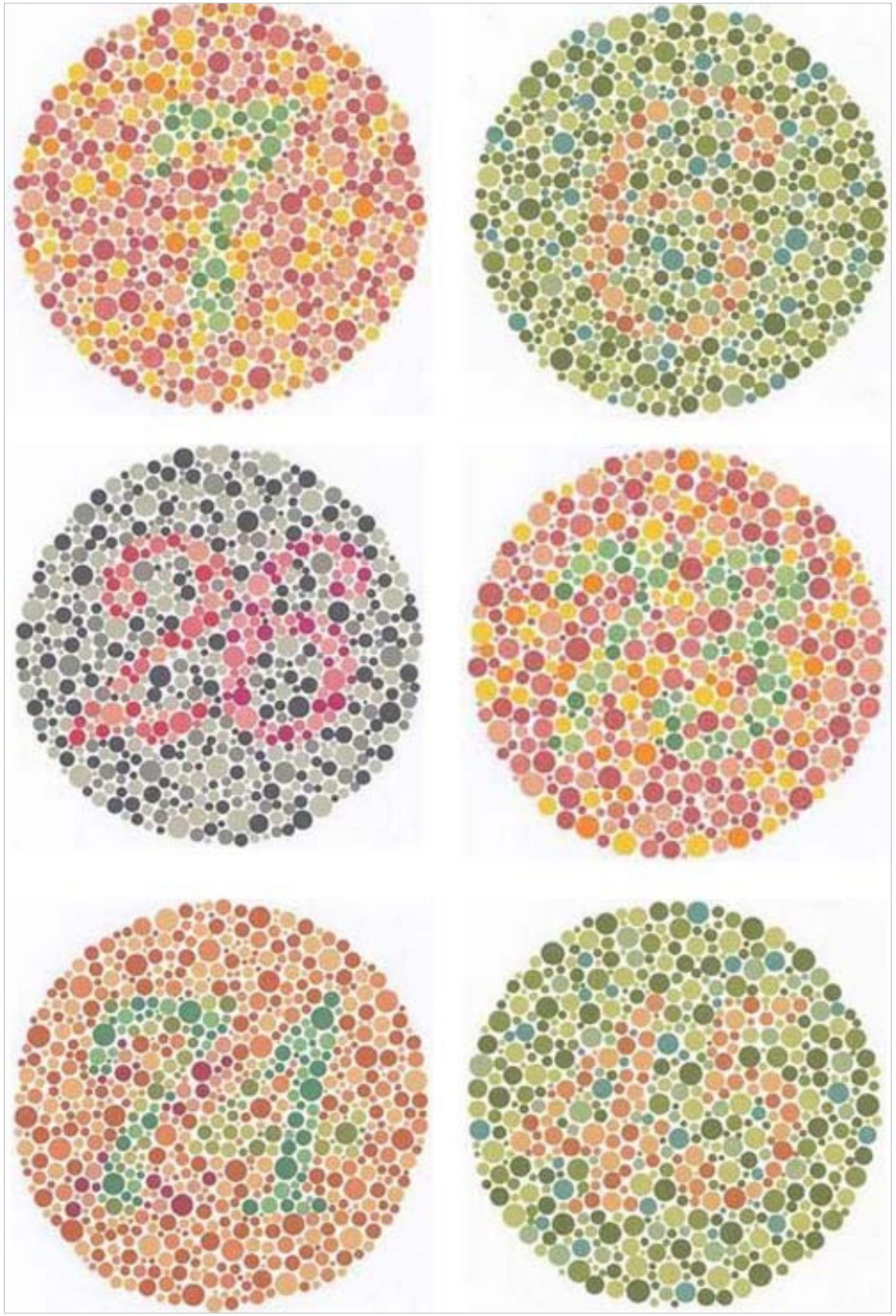
Using colour for continuous values



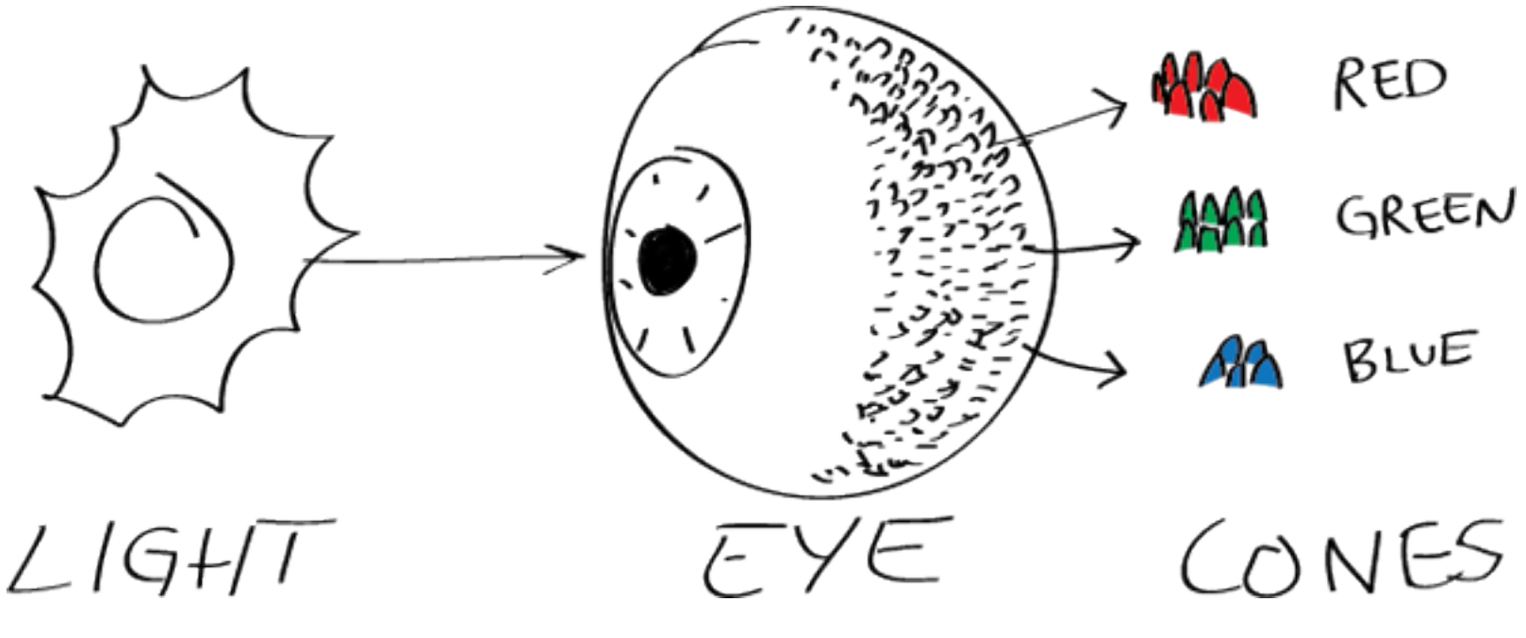


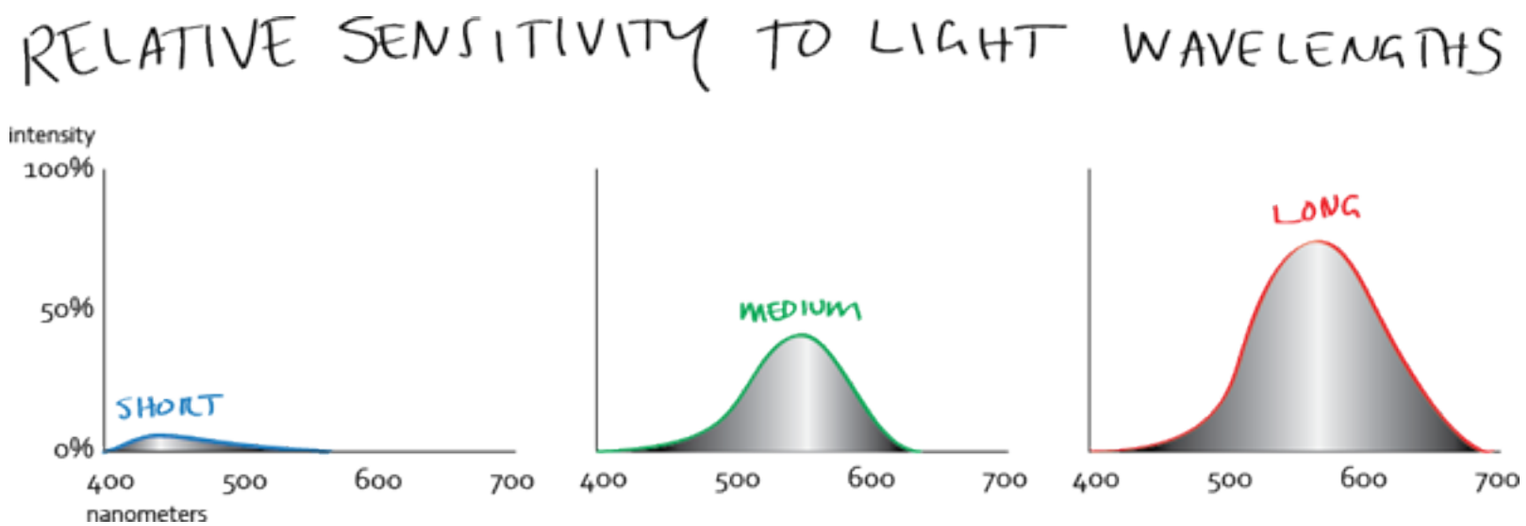
Protanopia affects 8% of males, 0.5% females of Northern European ancestry

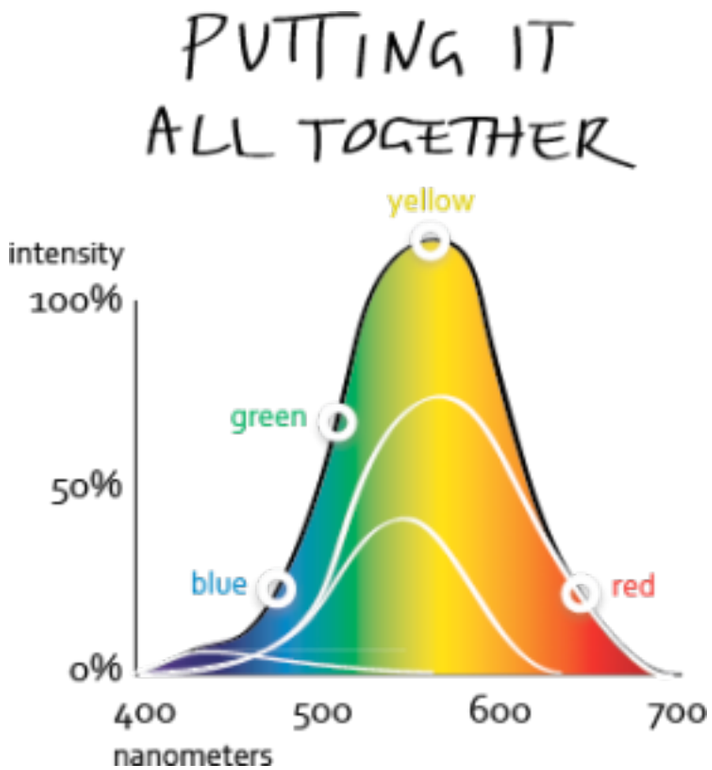
problem 2: colour sensitivity



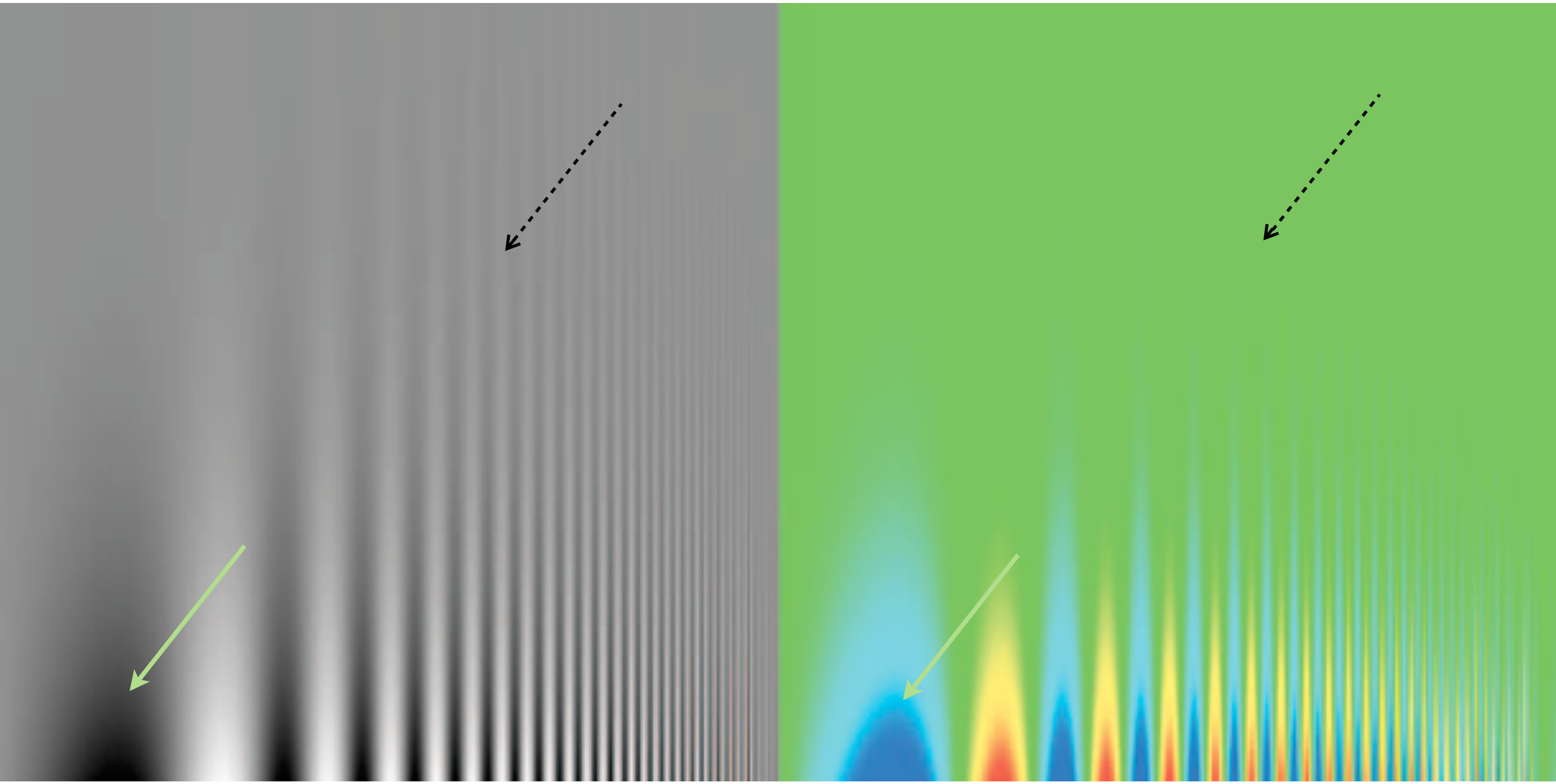
Using colour for continuous values problem 3: yellow is special





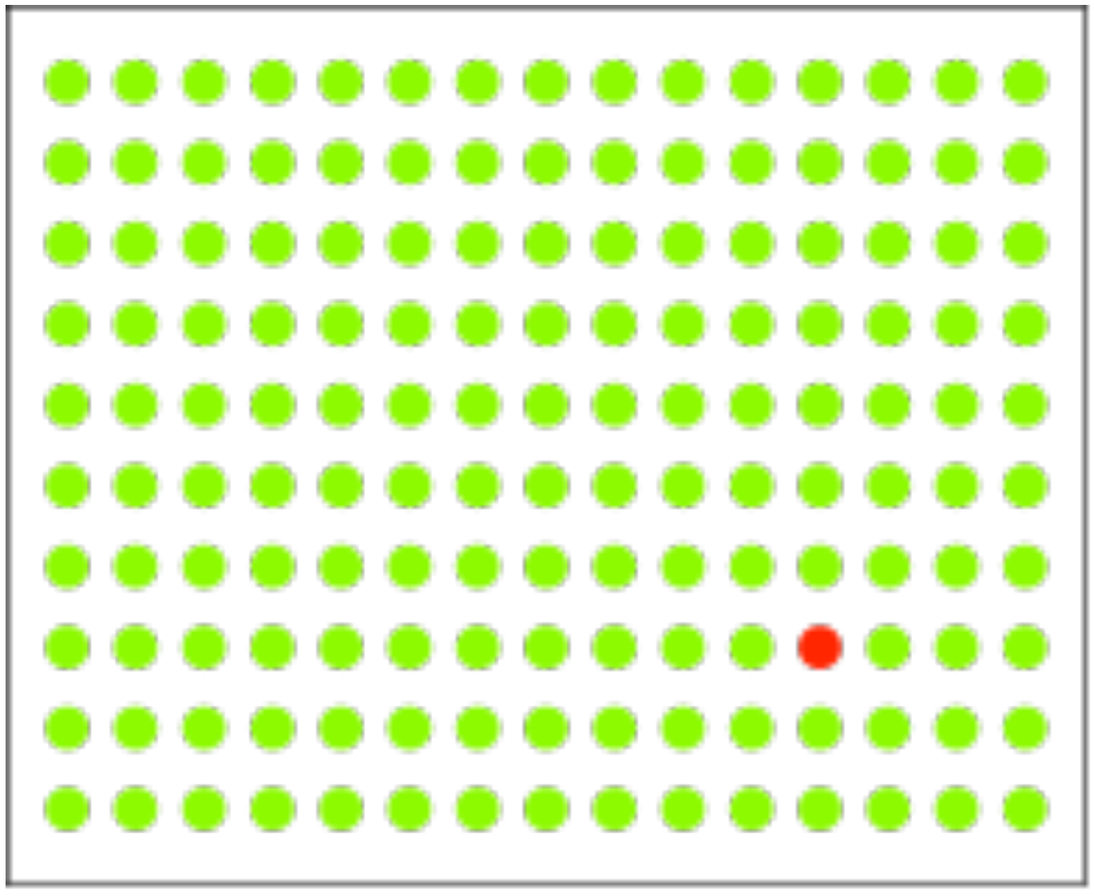


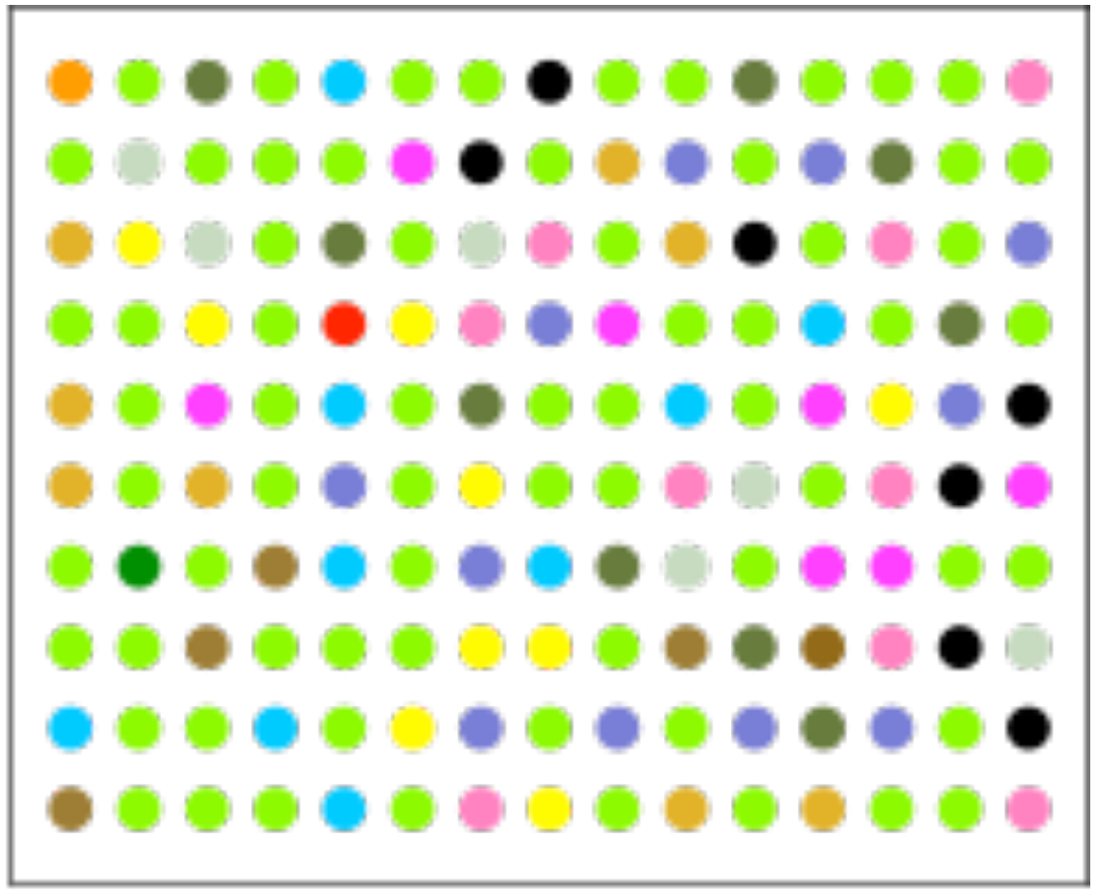
Using colour for continuous values
problem 4: Details: overemphasised or obscured

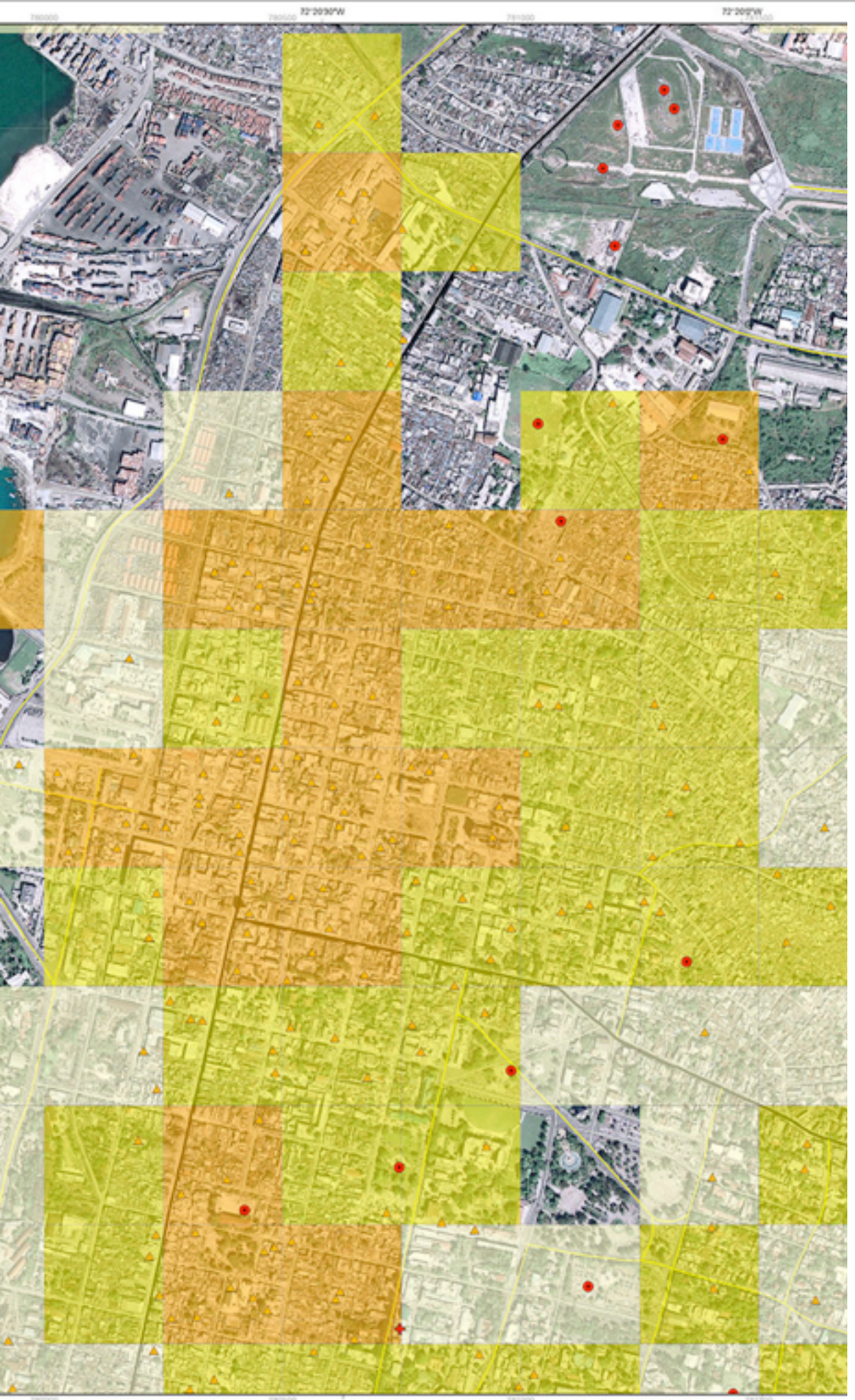


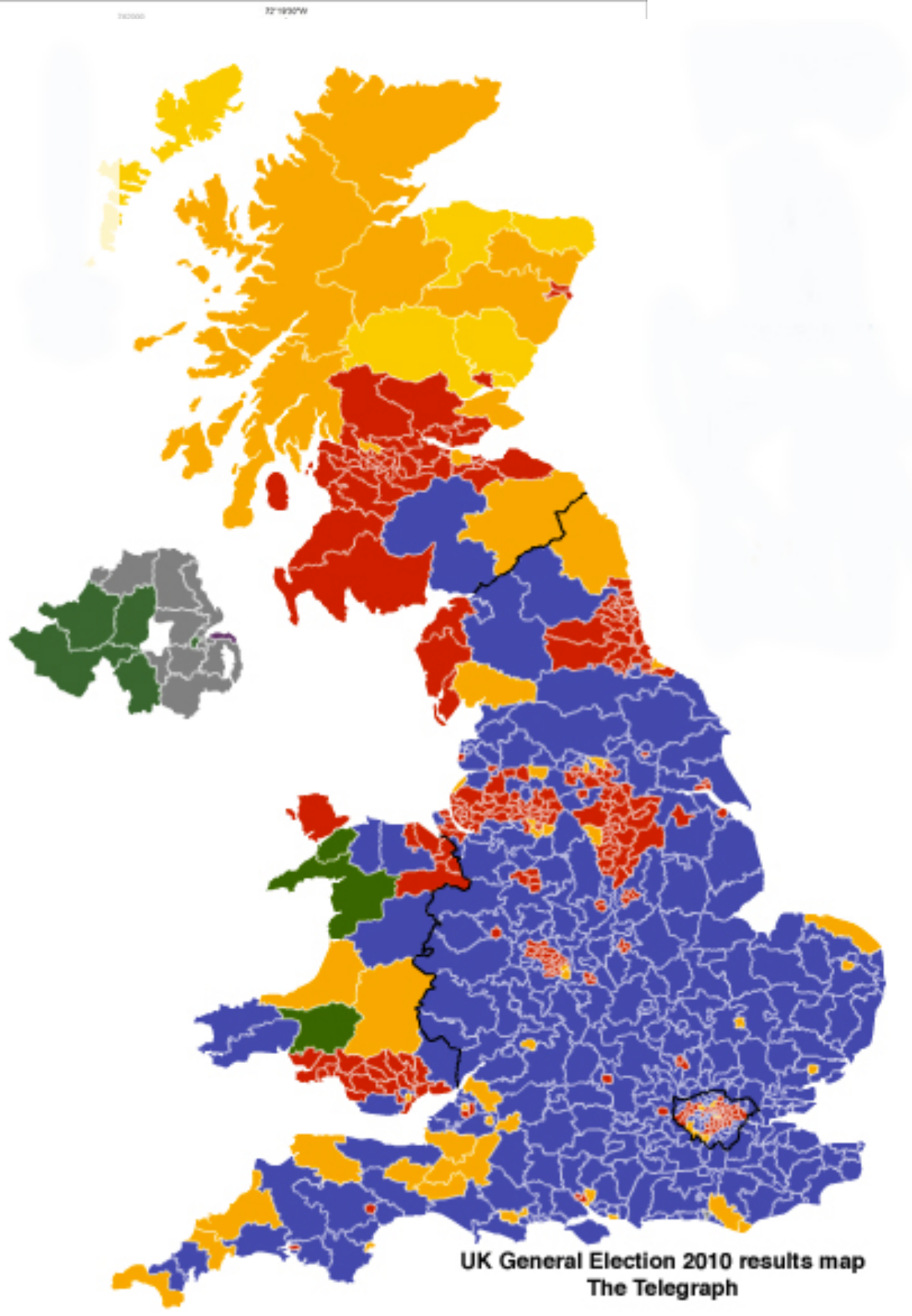
hue ‘borders’ overemphasise small changes, hue ‘middles’ blend potentially important details

Using colour for continuous values
problem 5: pop out can drown out



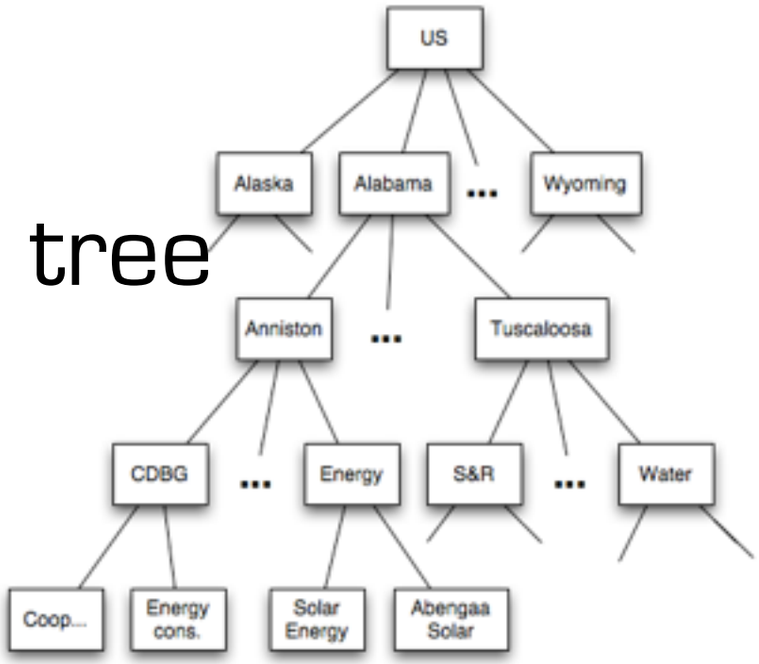


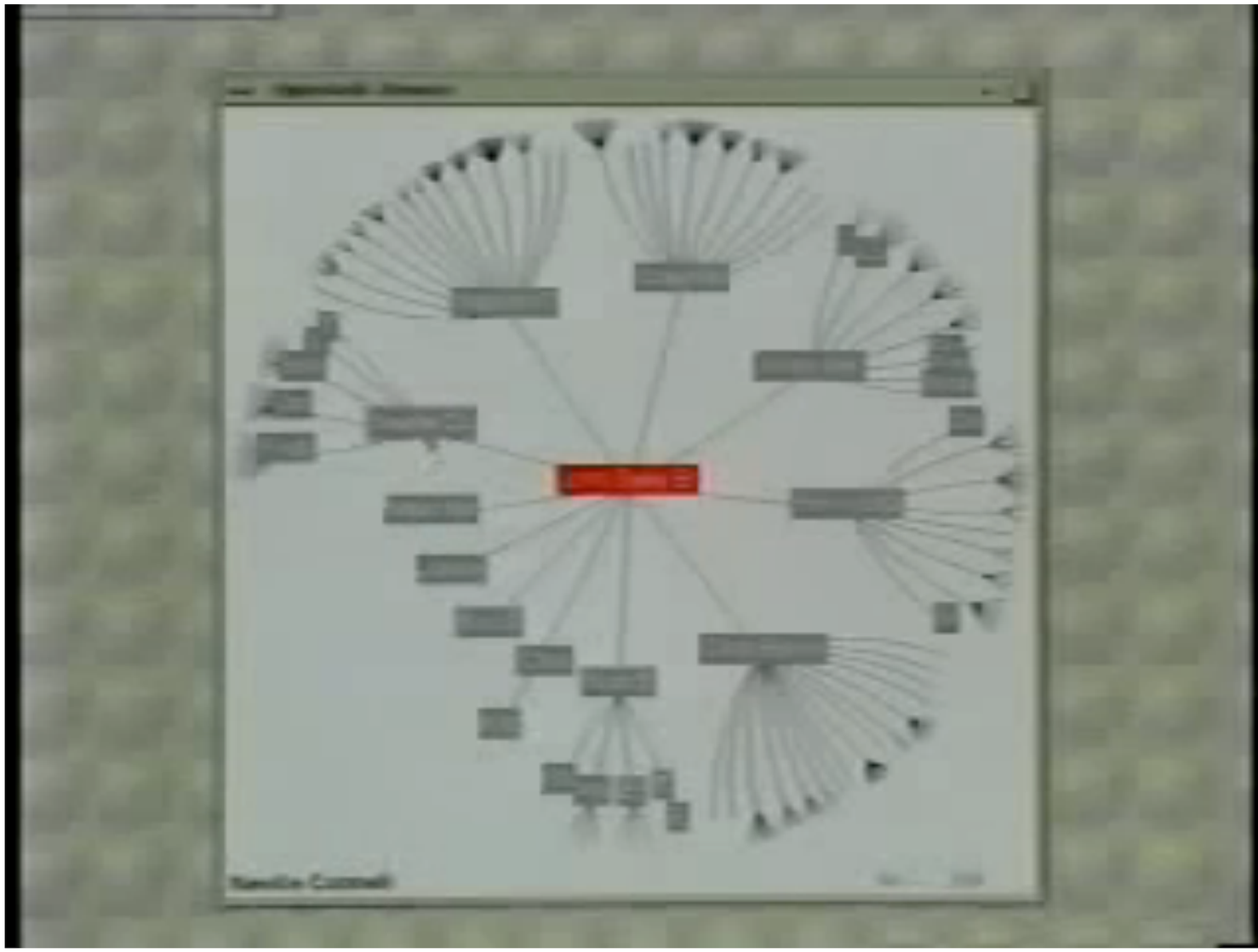


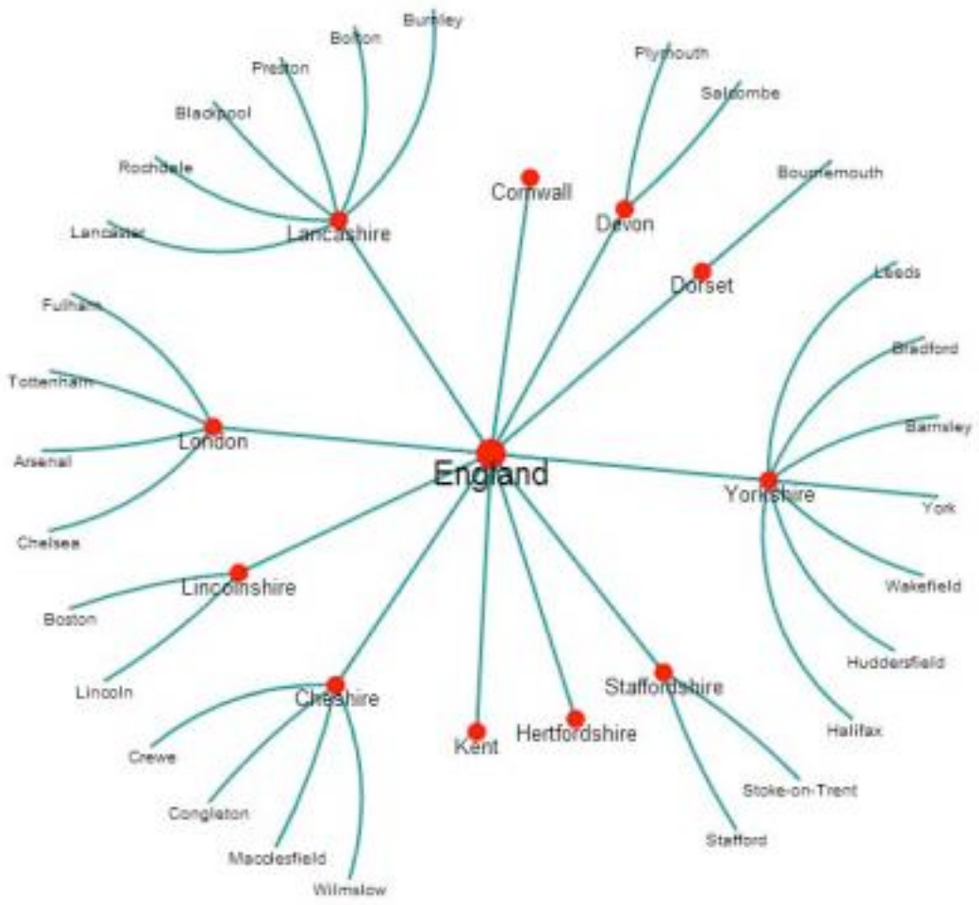

UK General Election 2010 results map
The Telegraph

multivariate relational data: hierarchical

tree



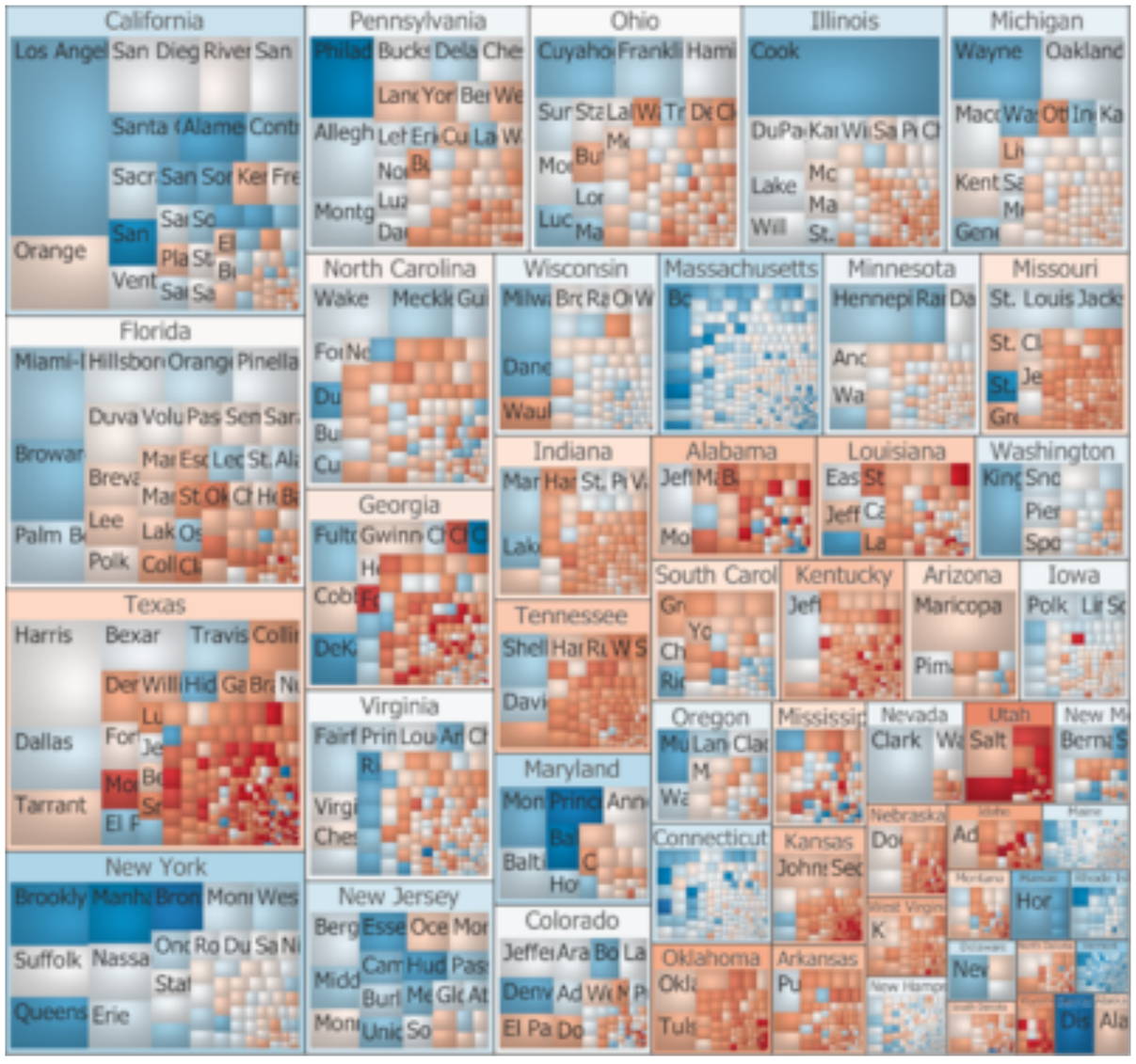


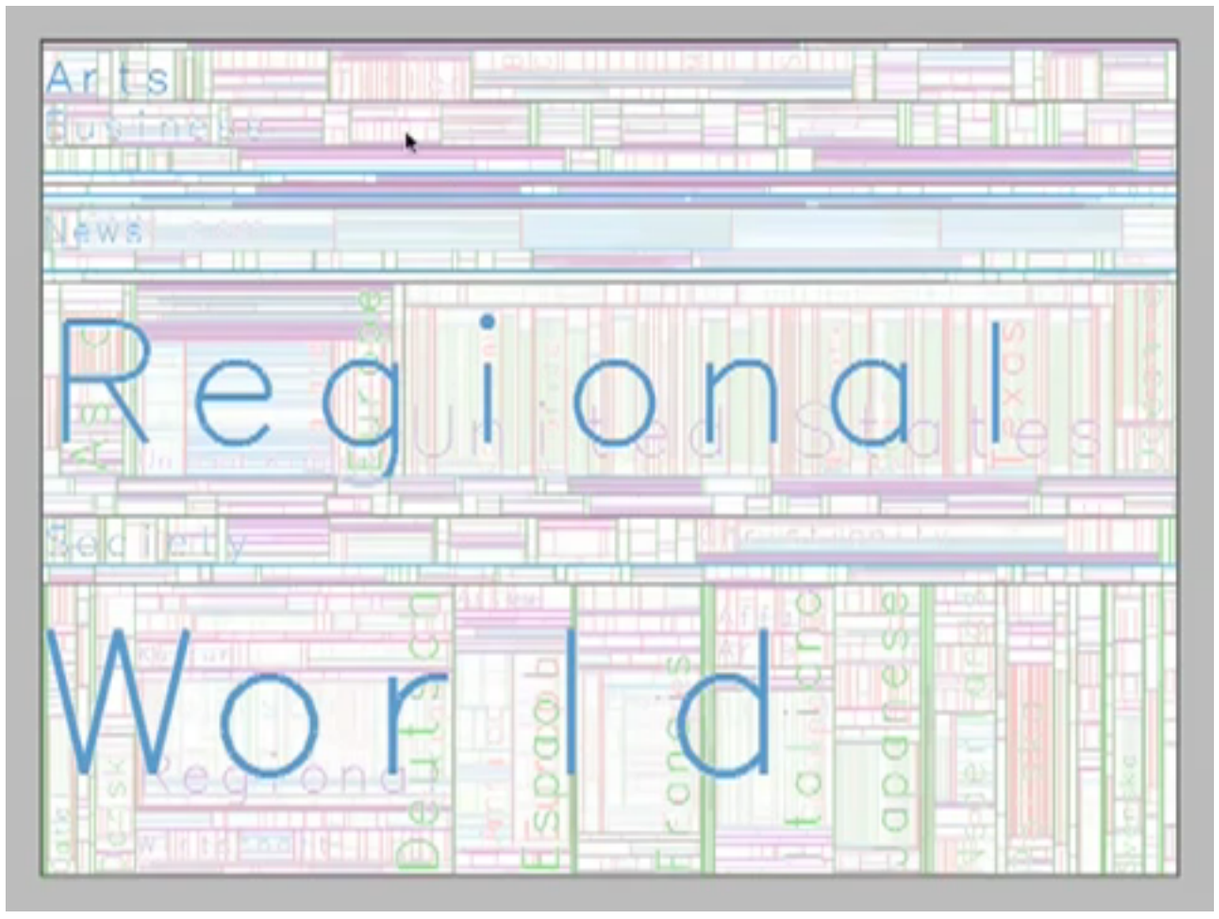


hyperbolic tree

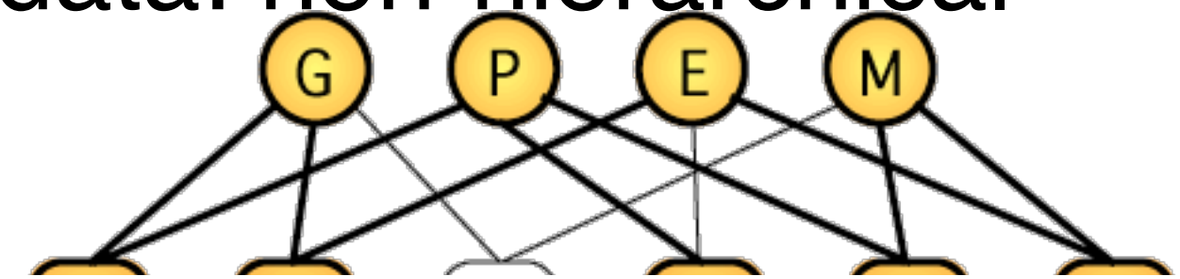
multivariate relational data: hierarchical

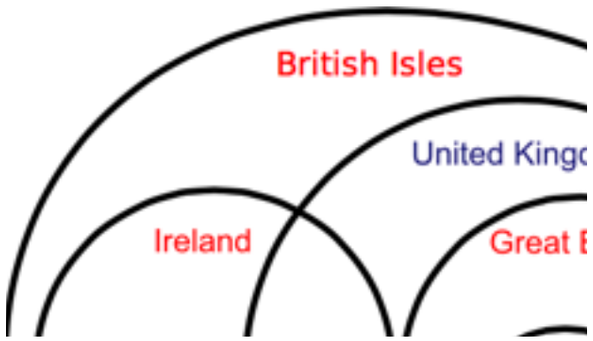
treemap



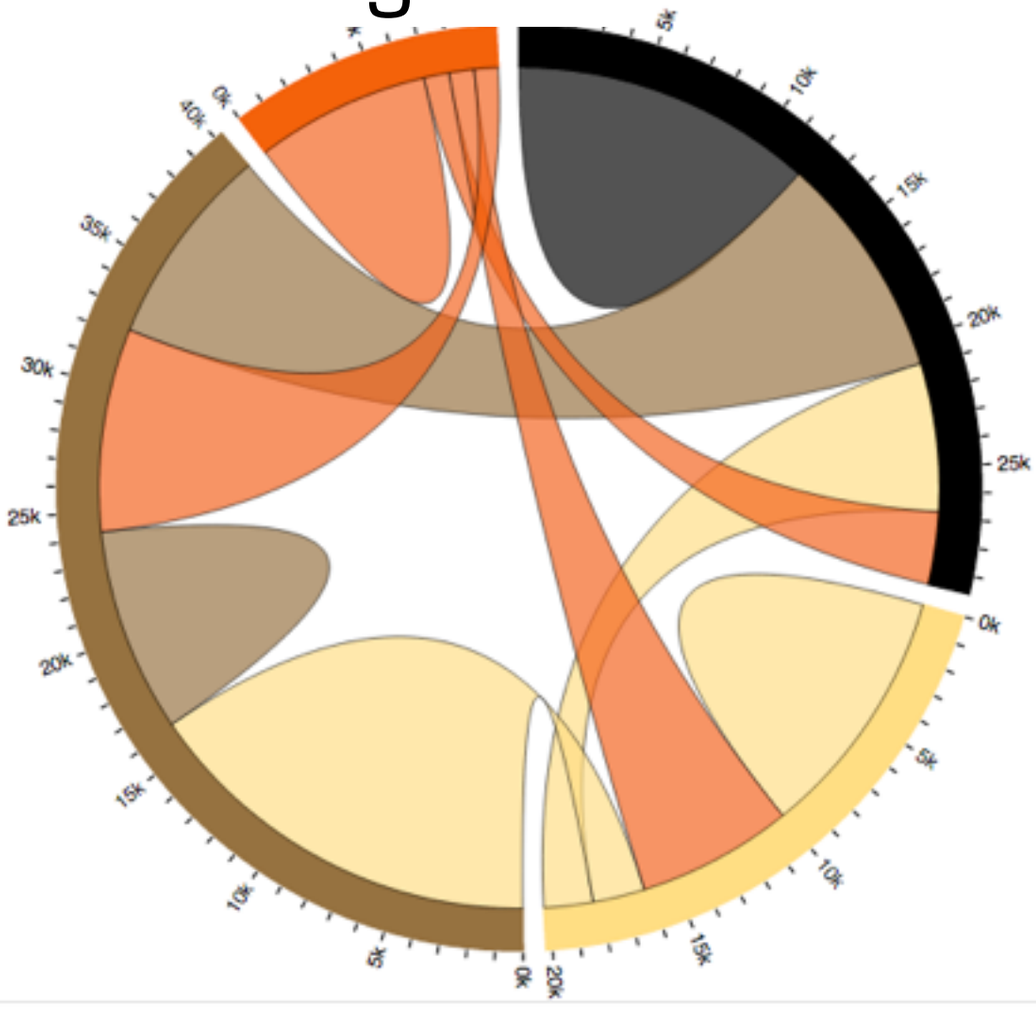


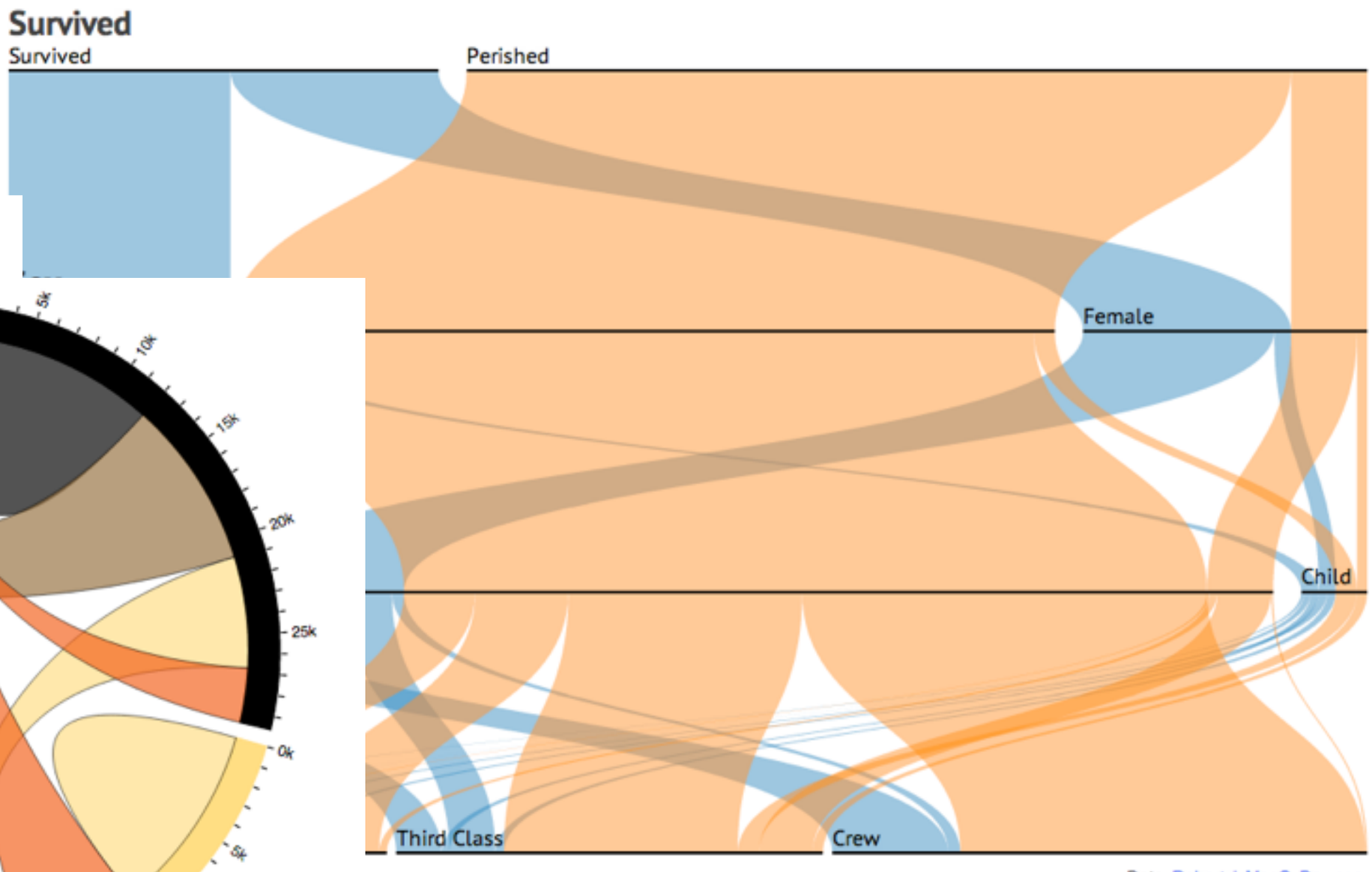
multivariate relational data: non-hierarchical





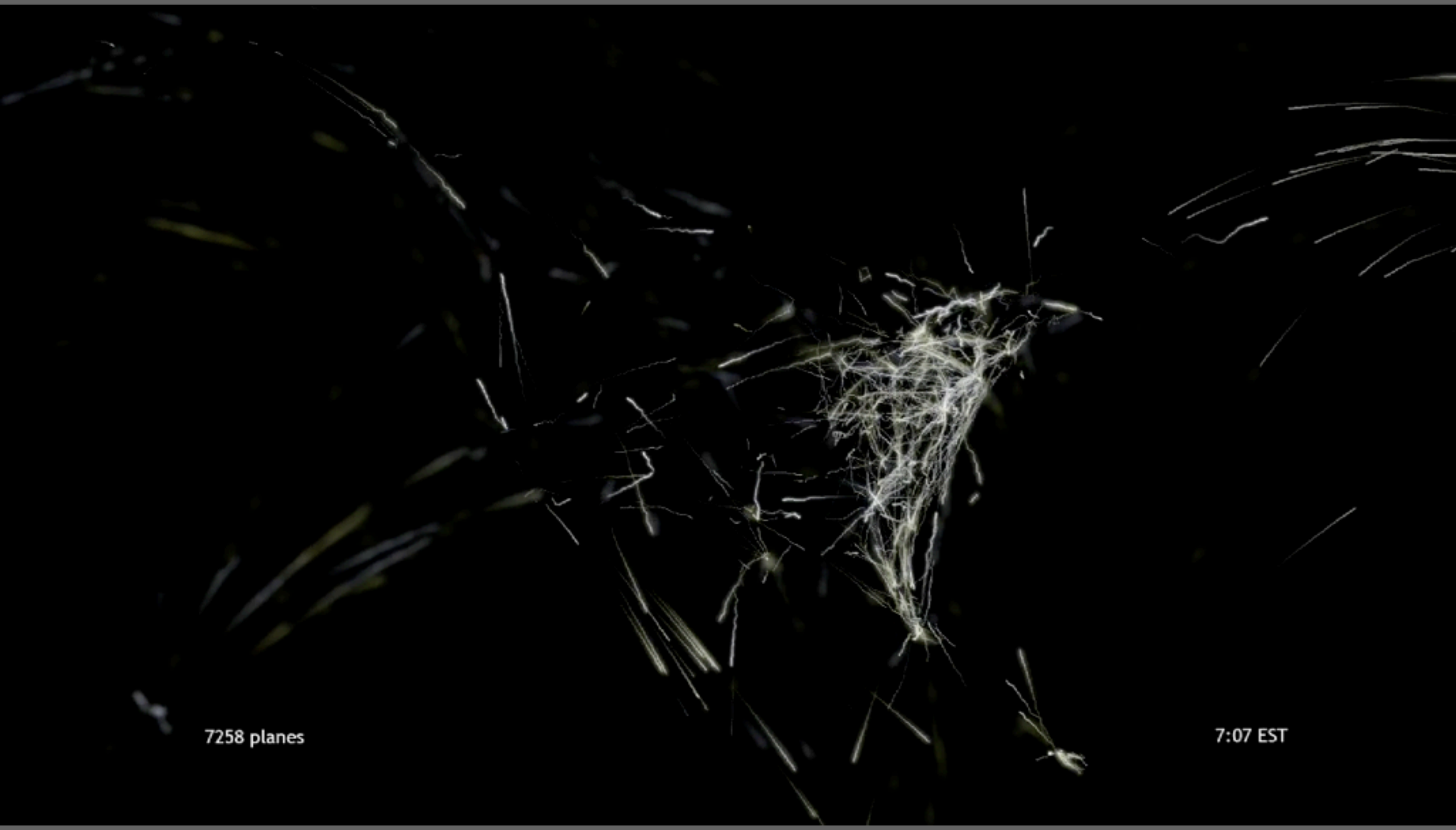
chord diagram





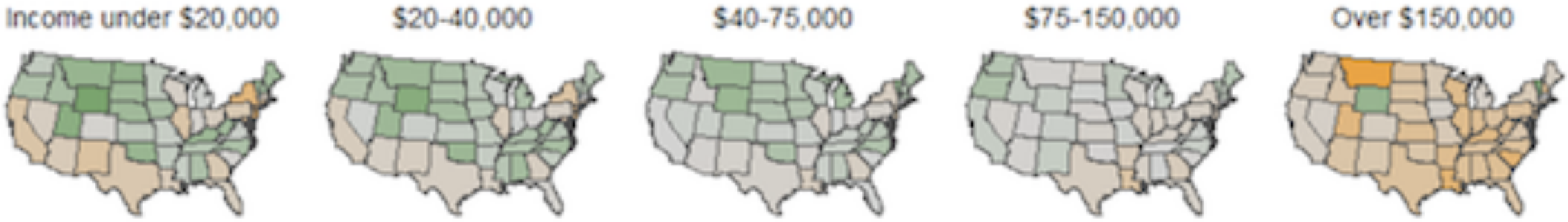
;S

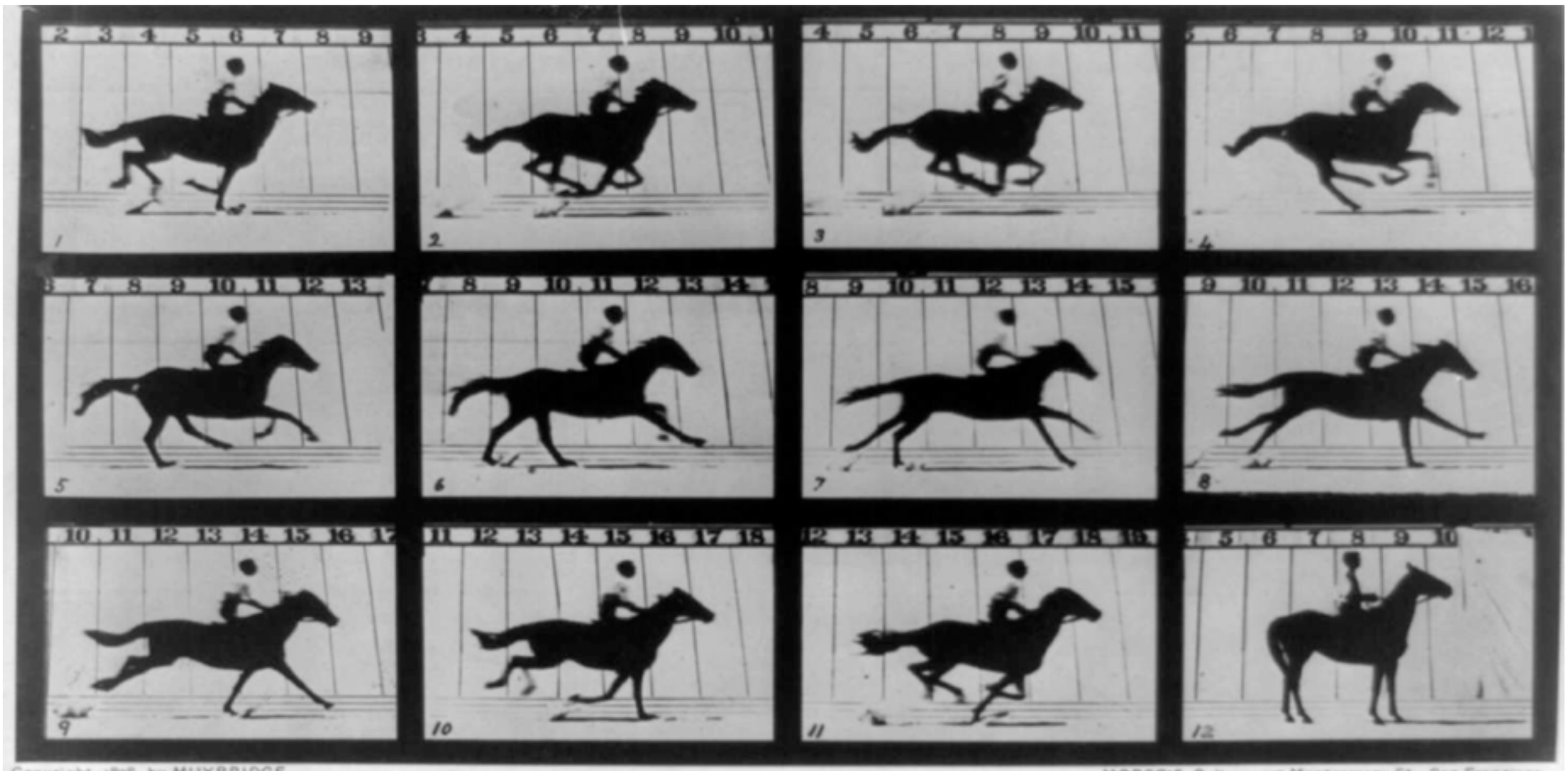
time series (animation)



aaron koblin - flight patterns

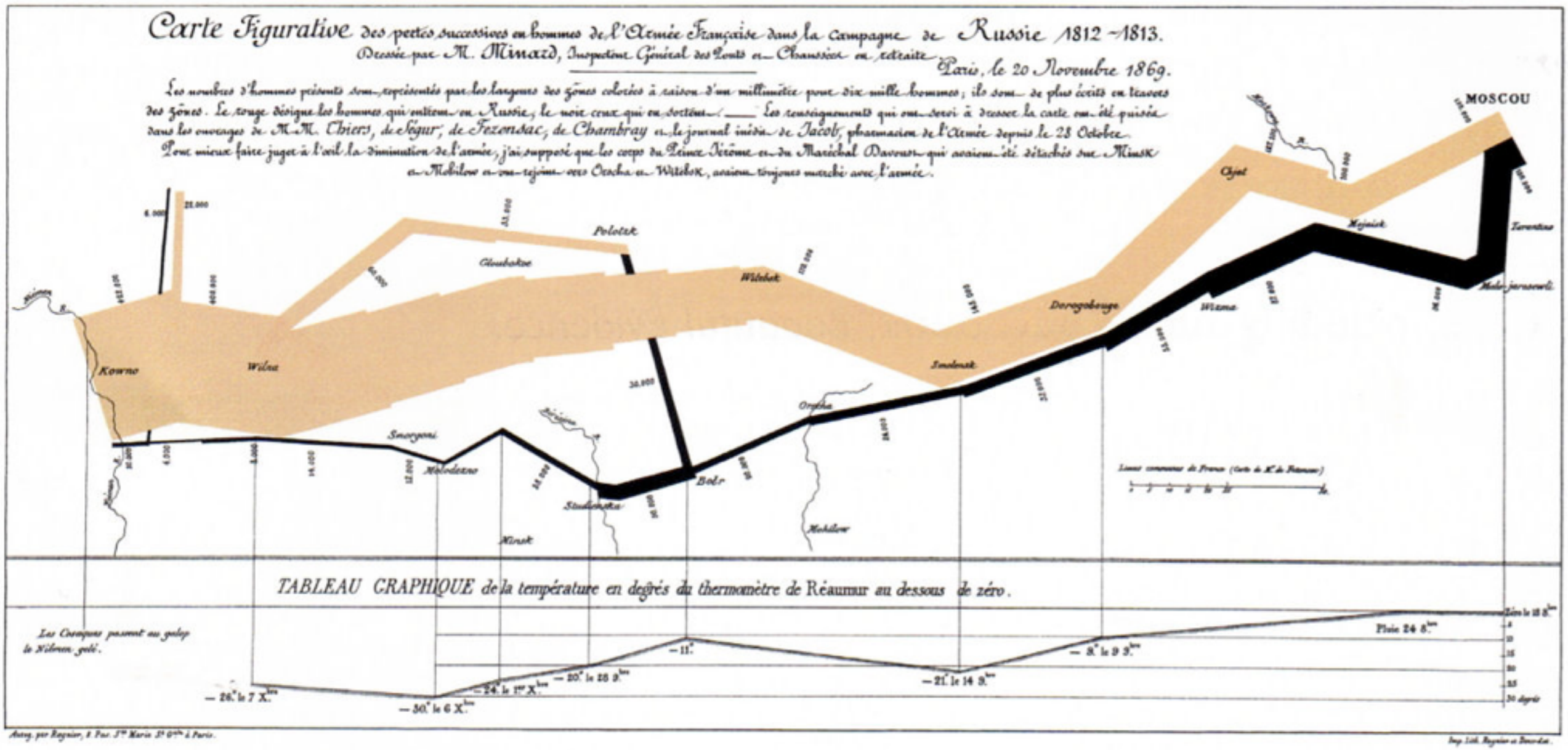
time series (static) - small multiples





charles joseph minard

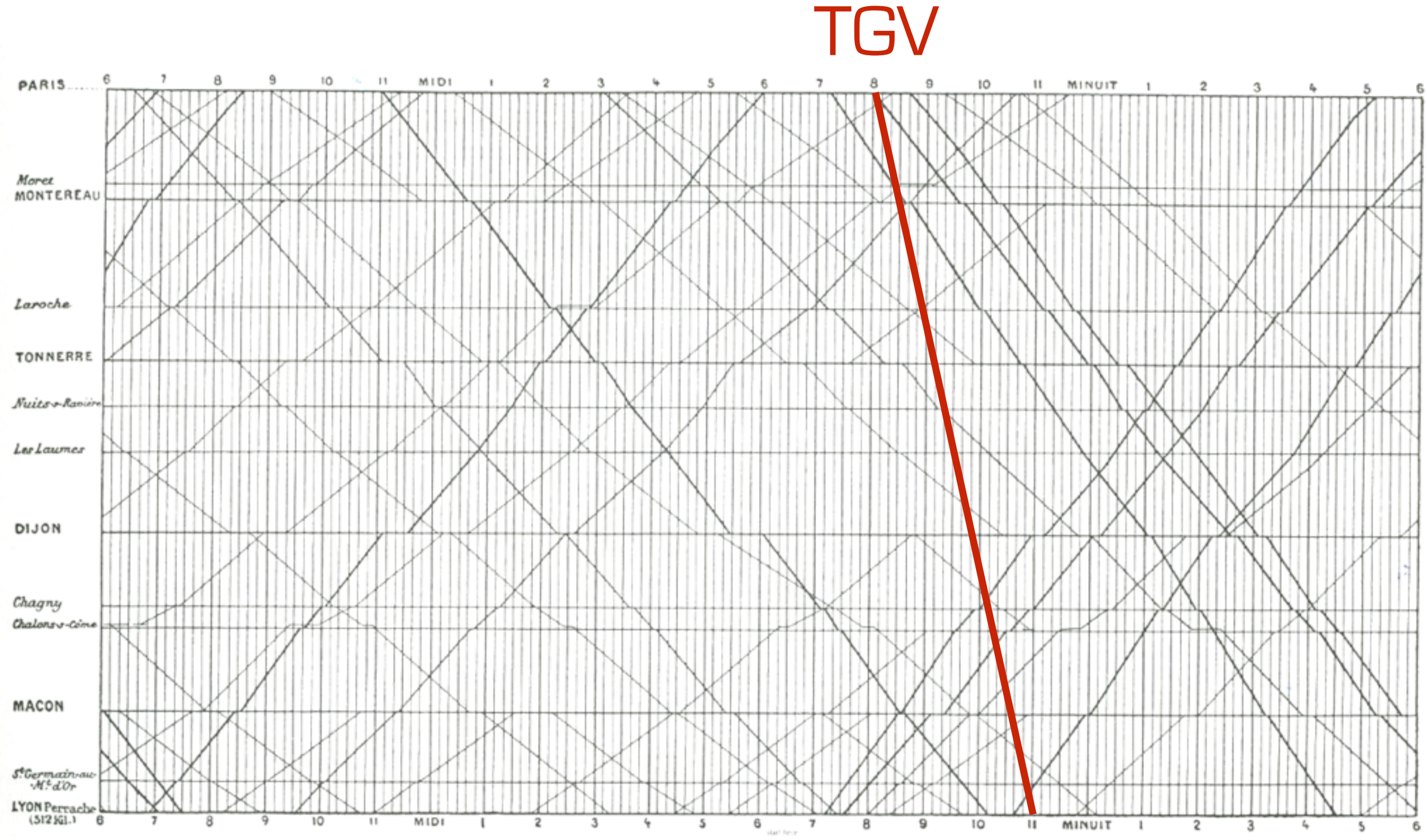
napoleon's march to moscow (1869)



multivariate

how many dimensions?

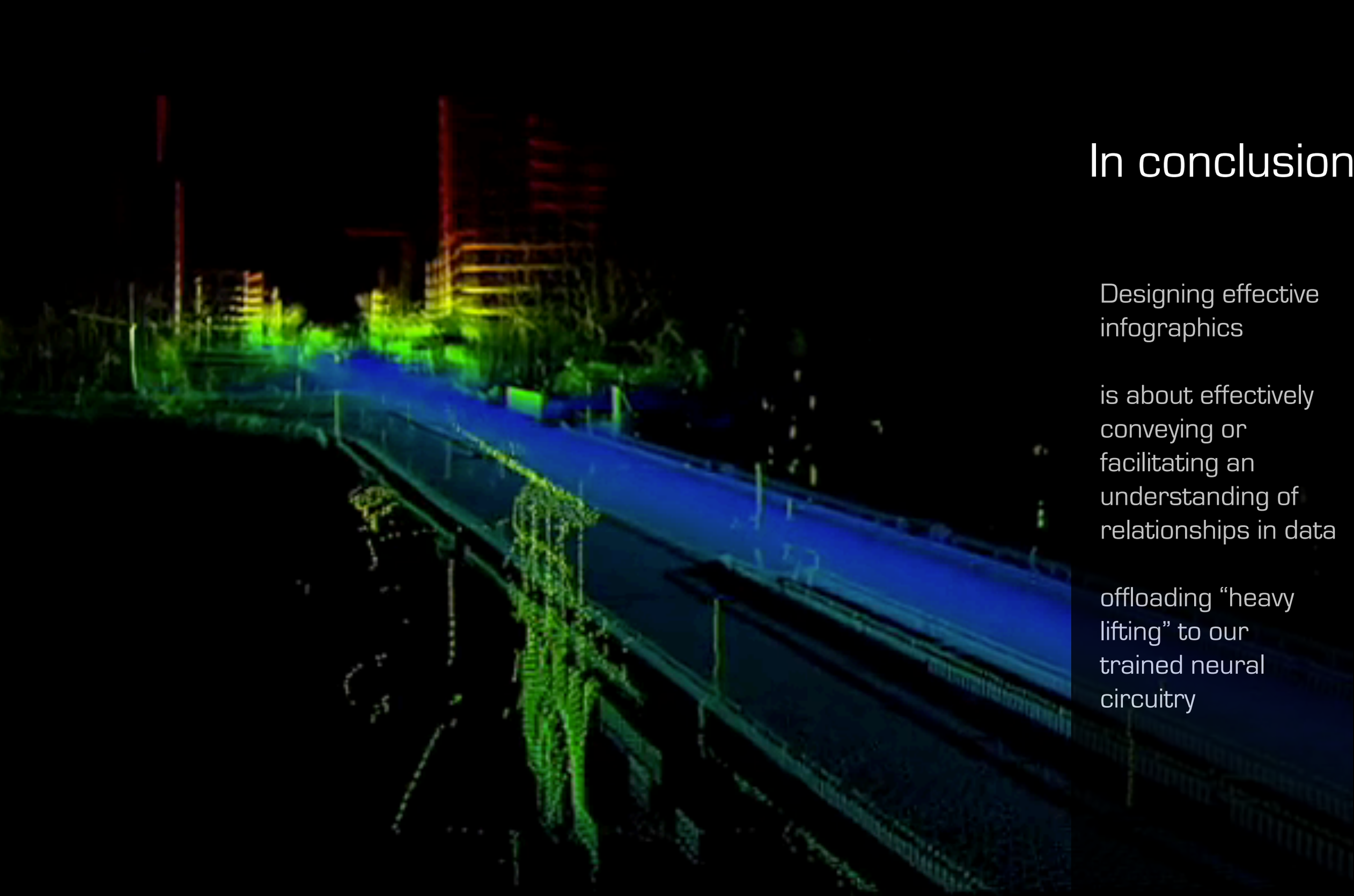
1) size of the army
2) advancing/retreating at each location
3) divisions
4) path taken by each
5) temperature
6) dates of waypoints



E.J. Marey
La méthode graphique
(1885)

200 years that changed the world

with Hans Rosling



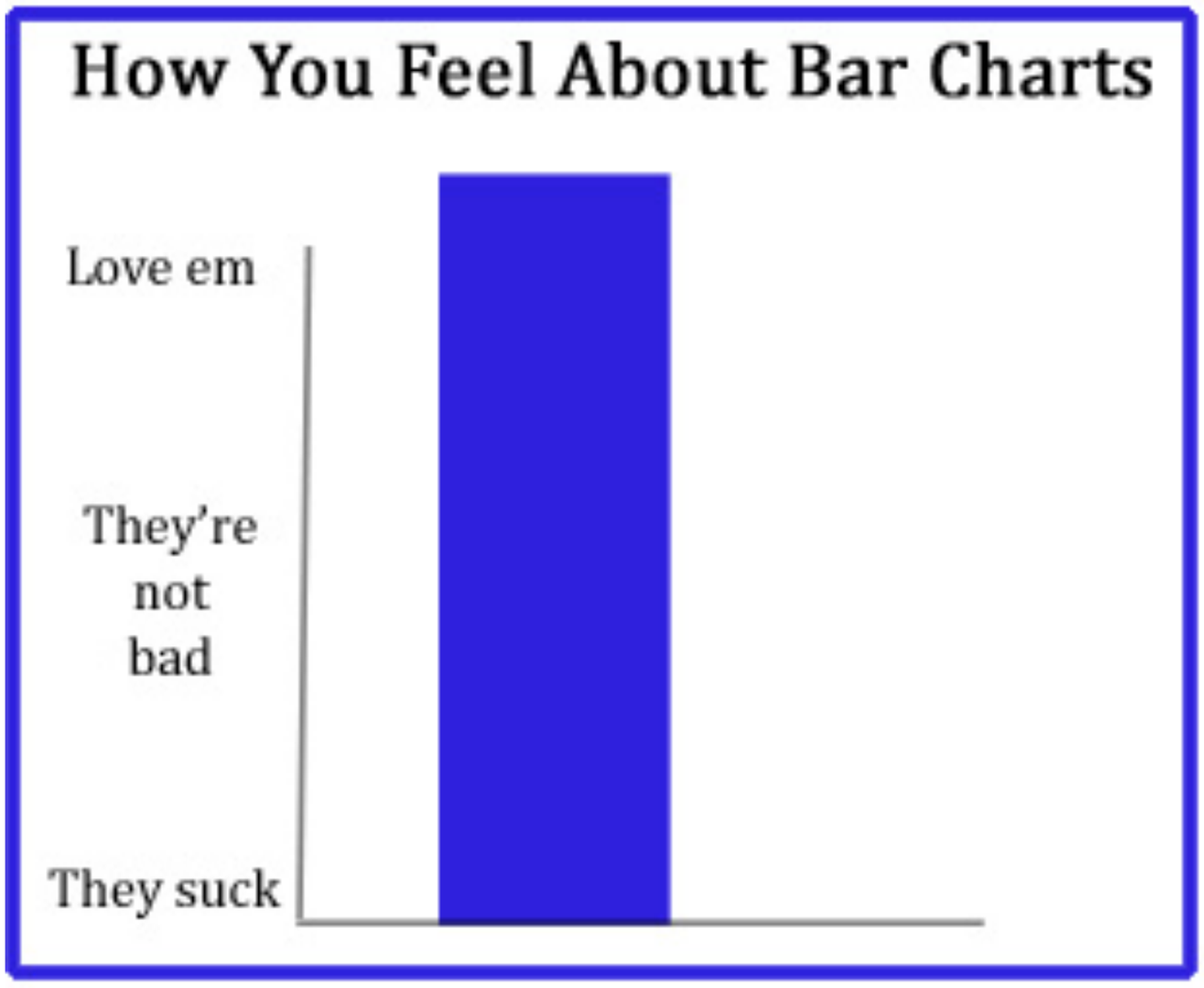
In conclusion

Designing effective infographics

is about effectively conveying or facilitating an understanding of relationships in data

offloading “heavy lifting” to our trained neural circuitry

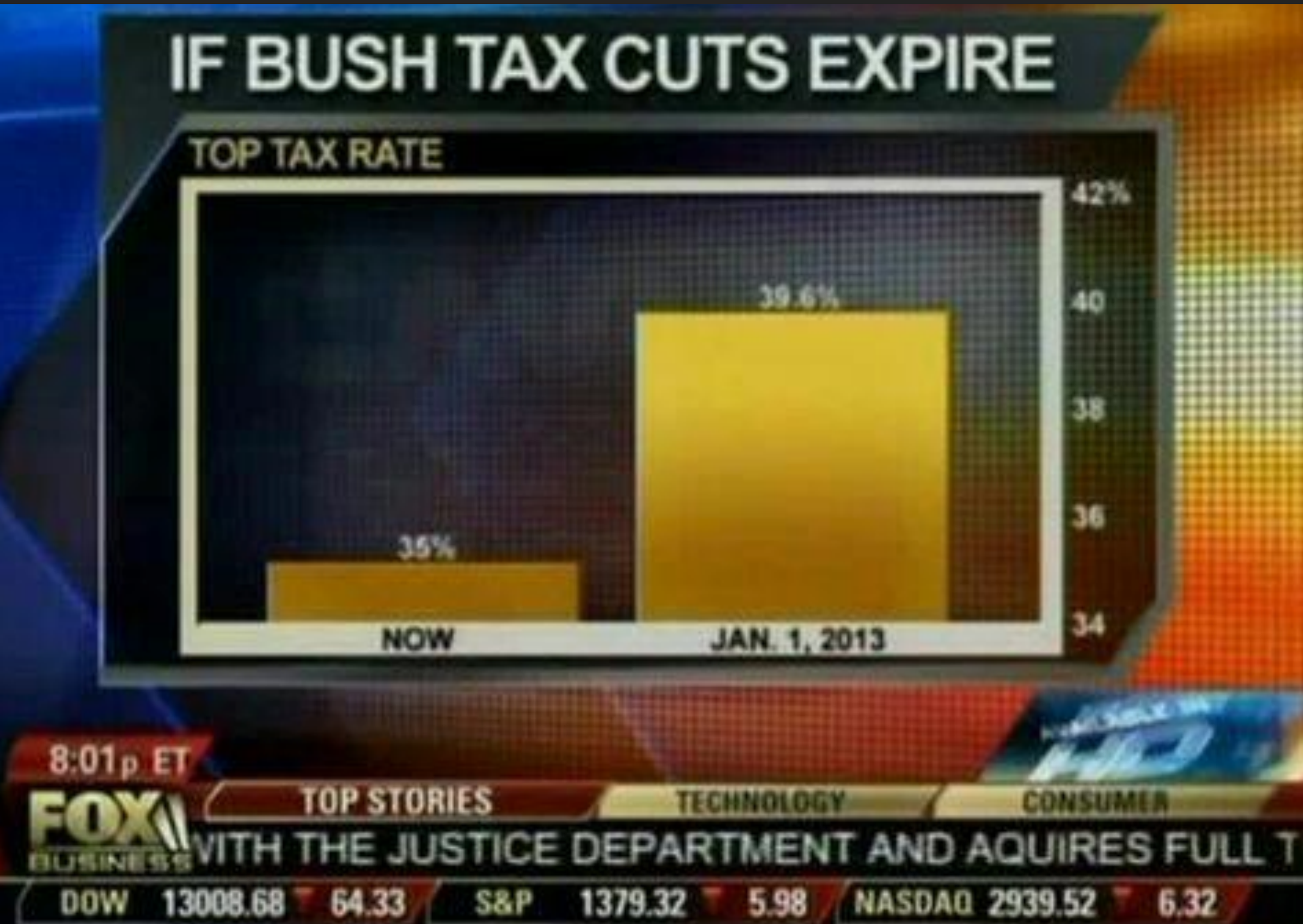
While still an art, many design principles grounded in usability can provide guidance: natural mappings, simplicity, & avoiding distortion



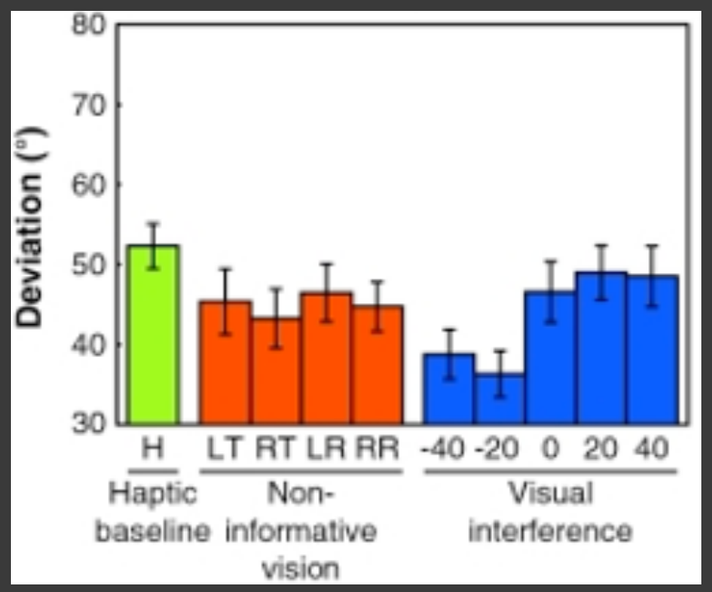
communicating through infographics:

visual + statistical sleight of hand to mislead the audience

1. Barchart baseline fail



1. Barchart baseline fail



Soaring gas prices

The price of a gallon of regular gas has risen 38 cents in South Florida in the past month. The national average rose 32 cents.

$4.0
3.9
3.8
3.7
3.6
$3.5
3.4
3.3
3.2
3.1
$3.0

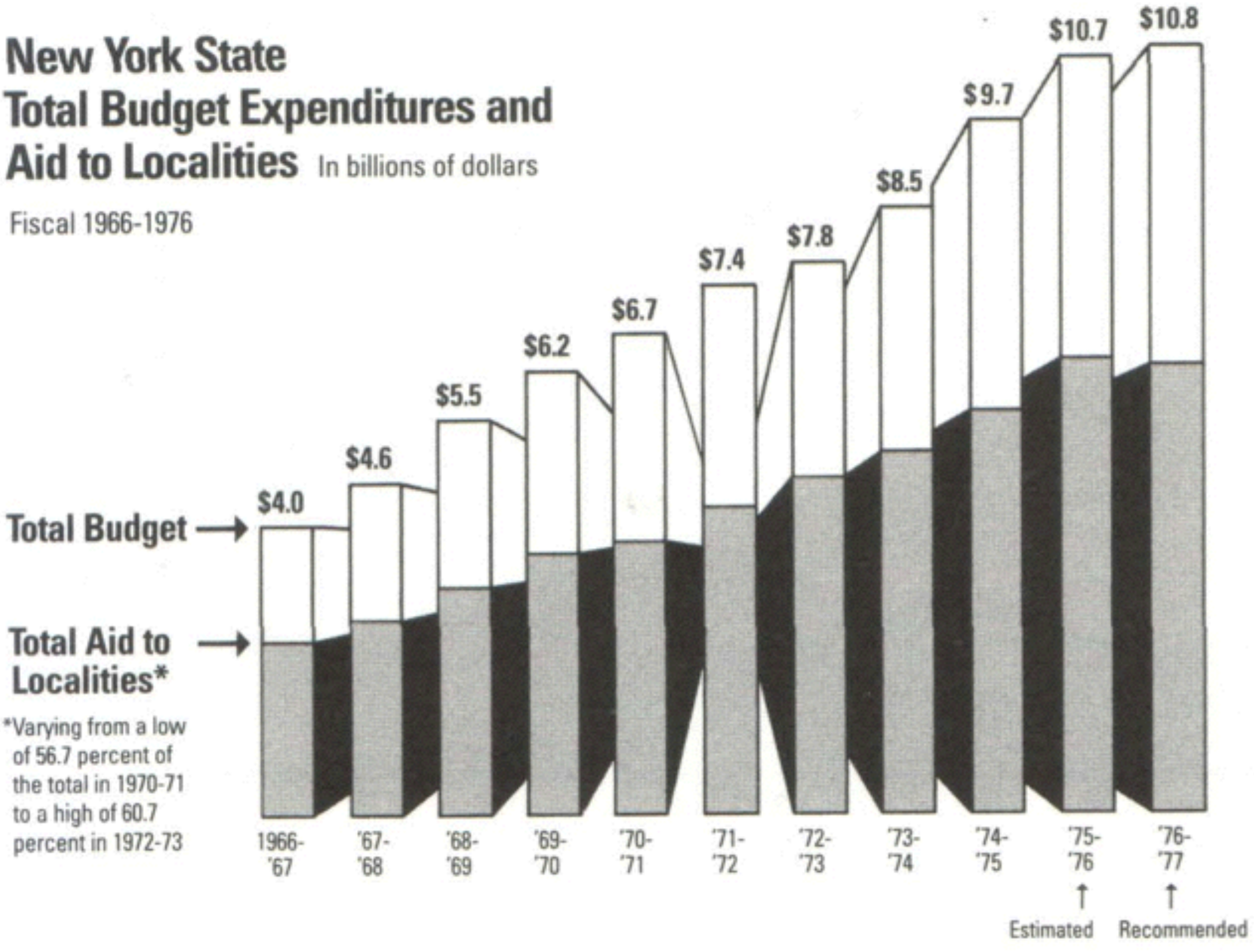
$3.565 Month Ago
$3.941 Now
$3.529 Month Ago
$3.918 Now
$3.432 Month Ago
$3.789 Now
$3.50 Month Ago
$3.826 Now

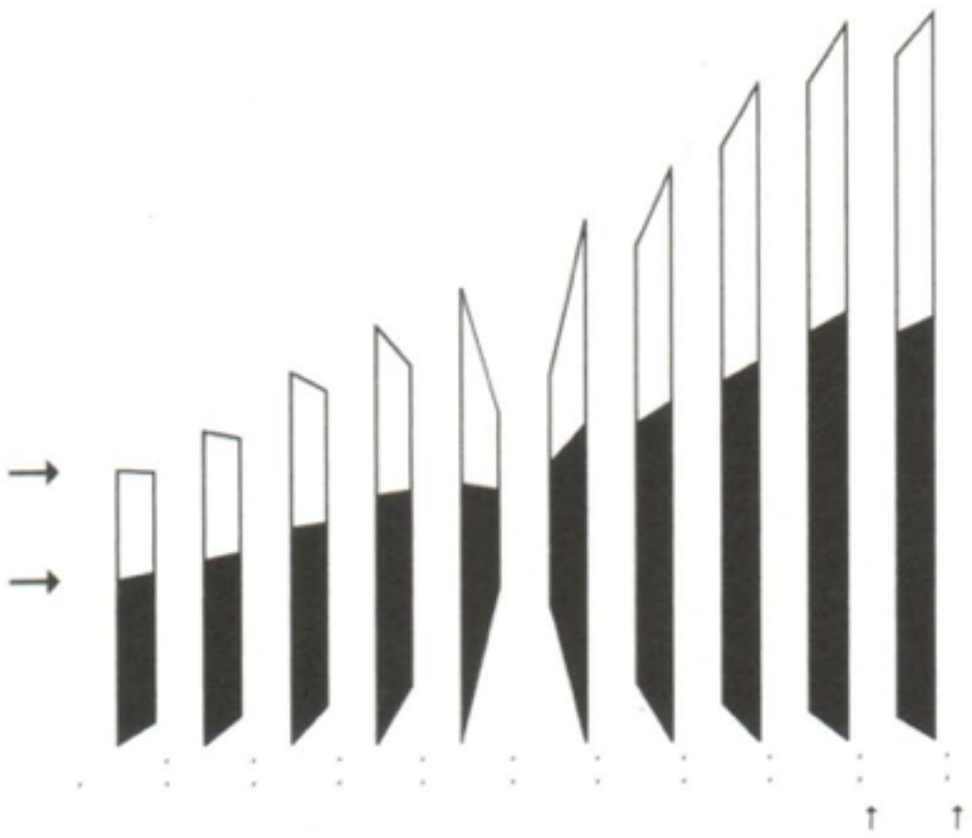
Miami
Ft. Lauderdale
Florida
U.S.

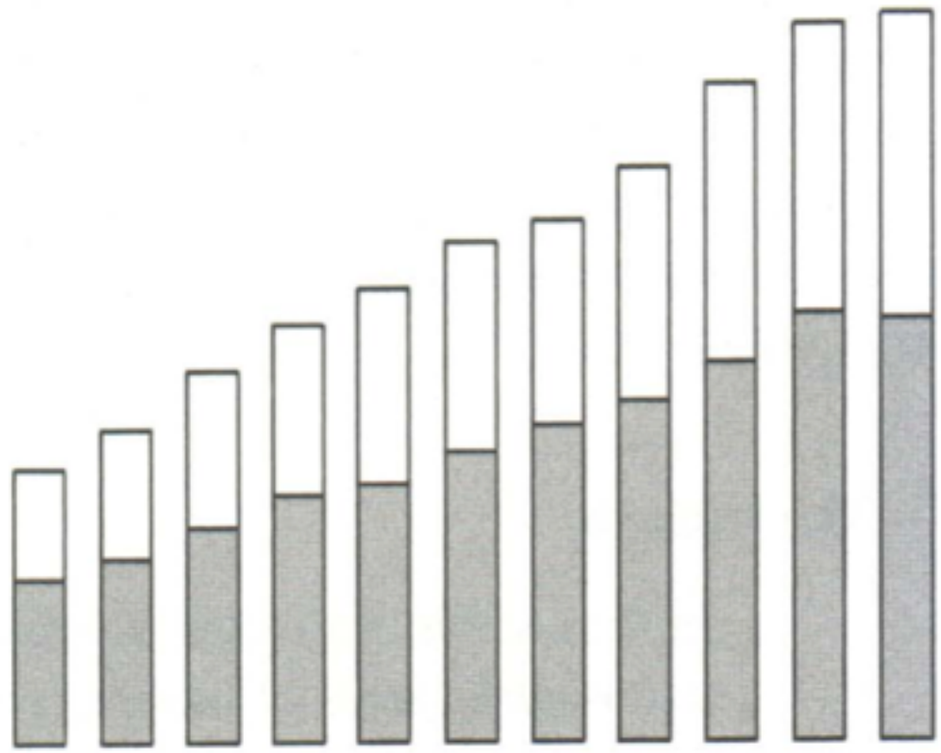
Source: AAA Fuel Gauge Report

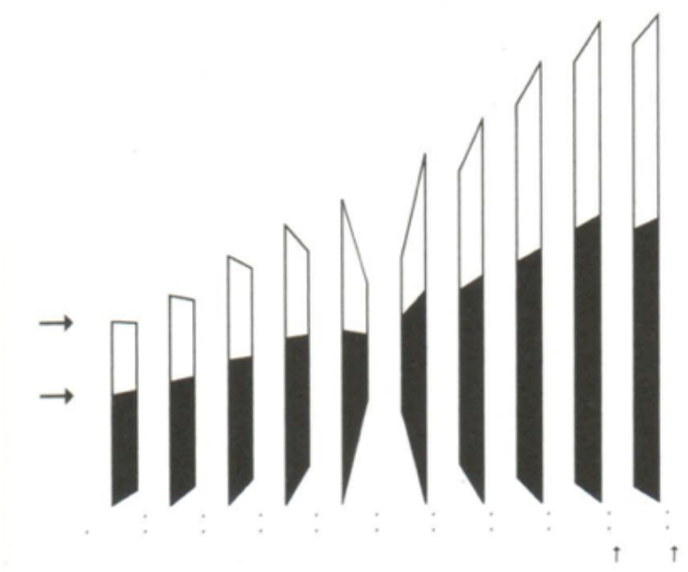
THE MIAMI HERALD

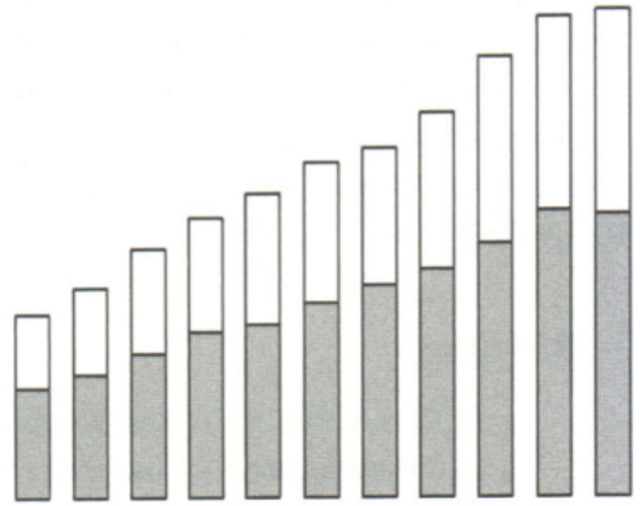
2. Perspective and measurement fail



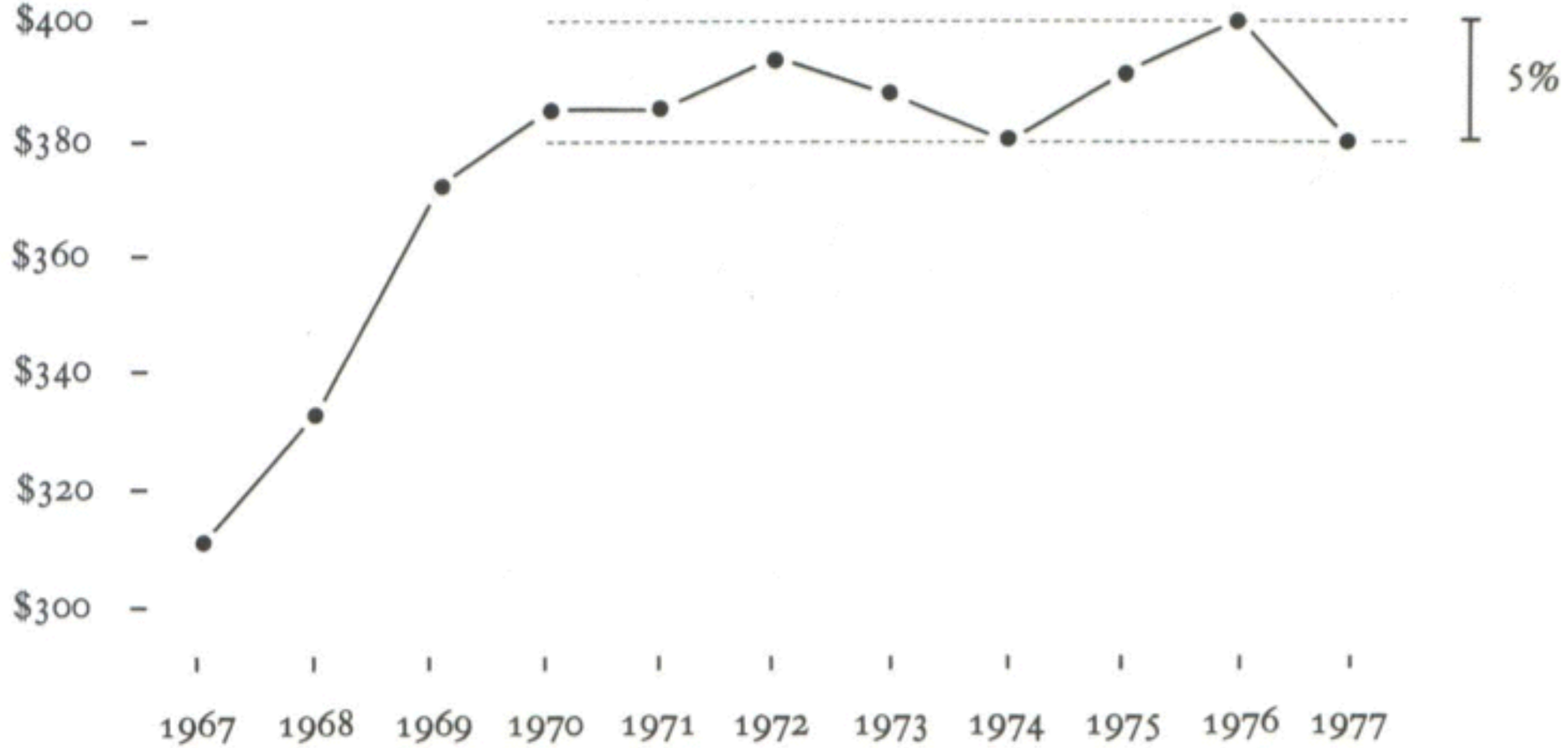




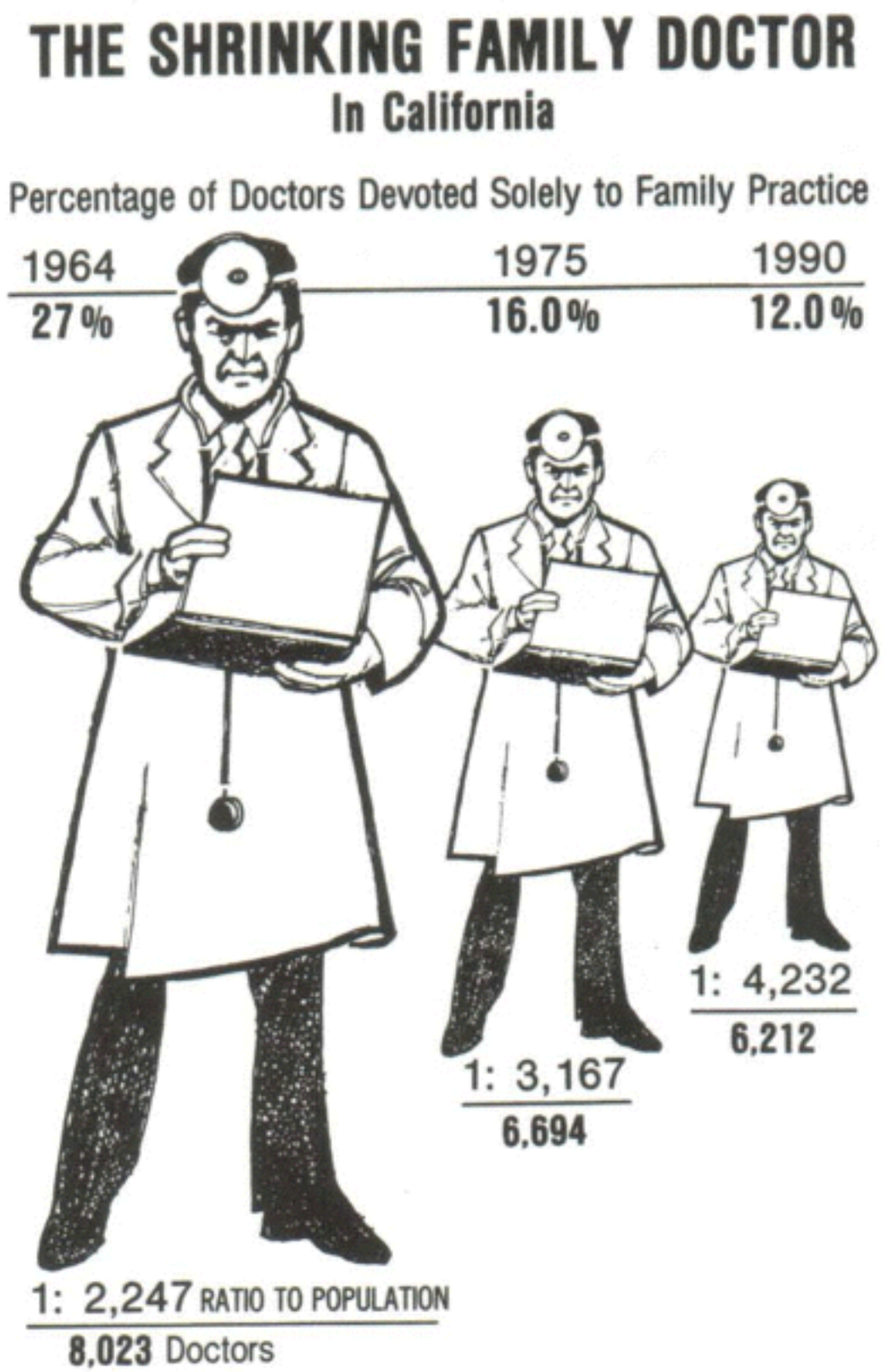


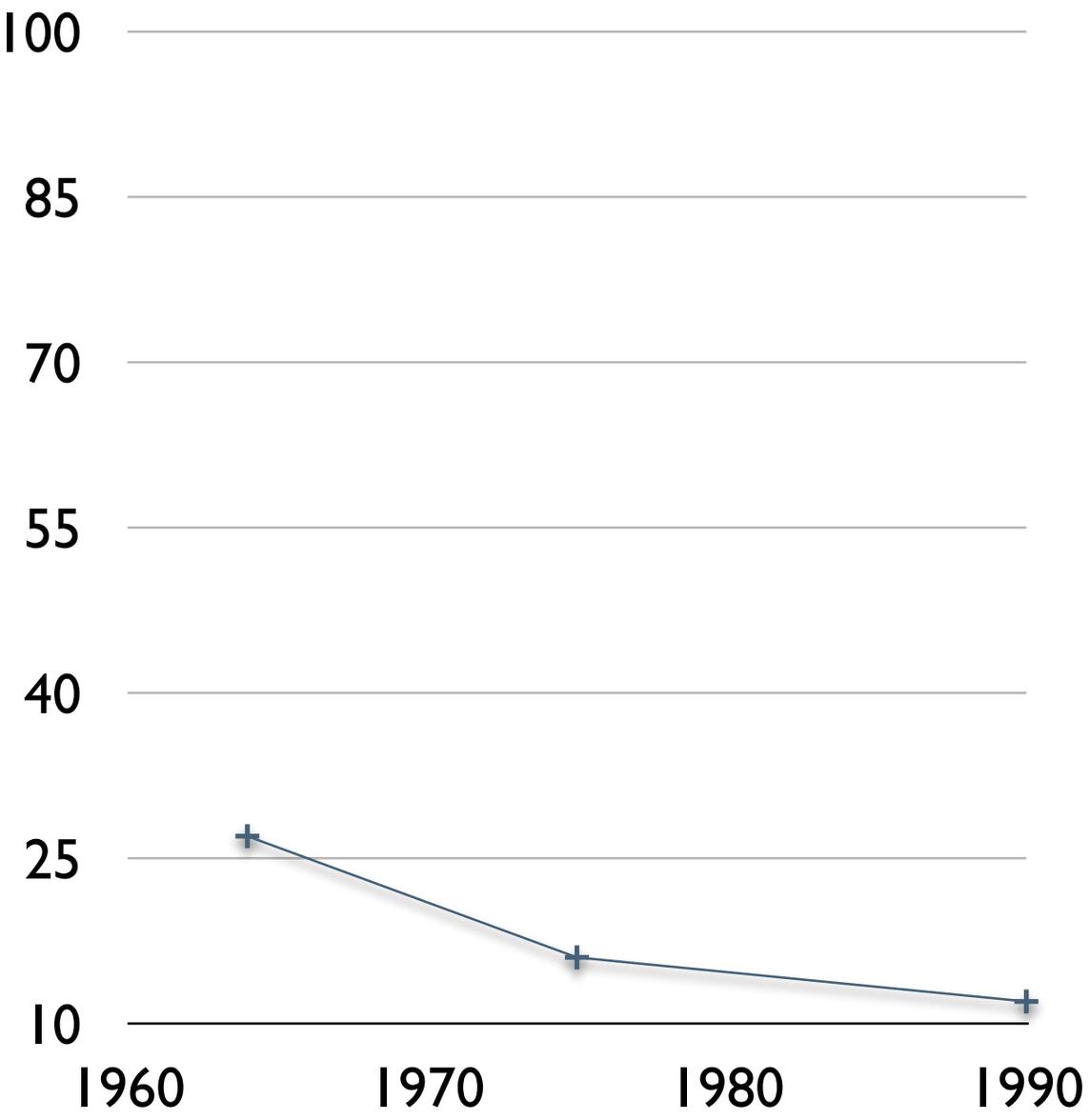


Per capita
budget expenditures,
in constant dollars

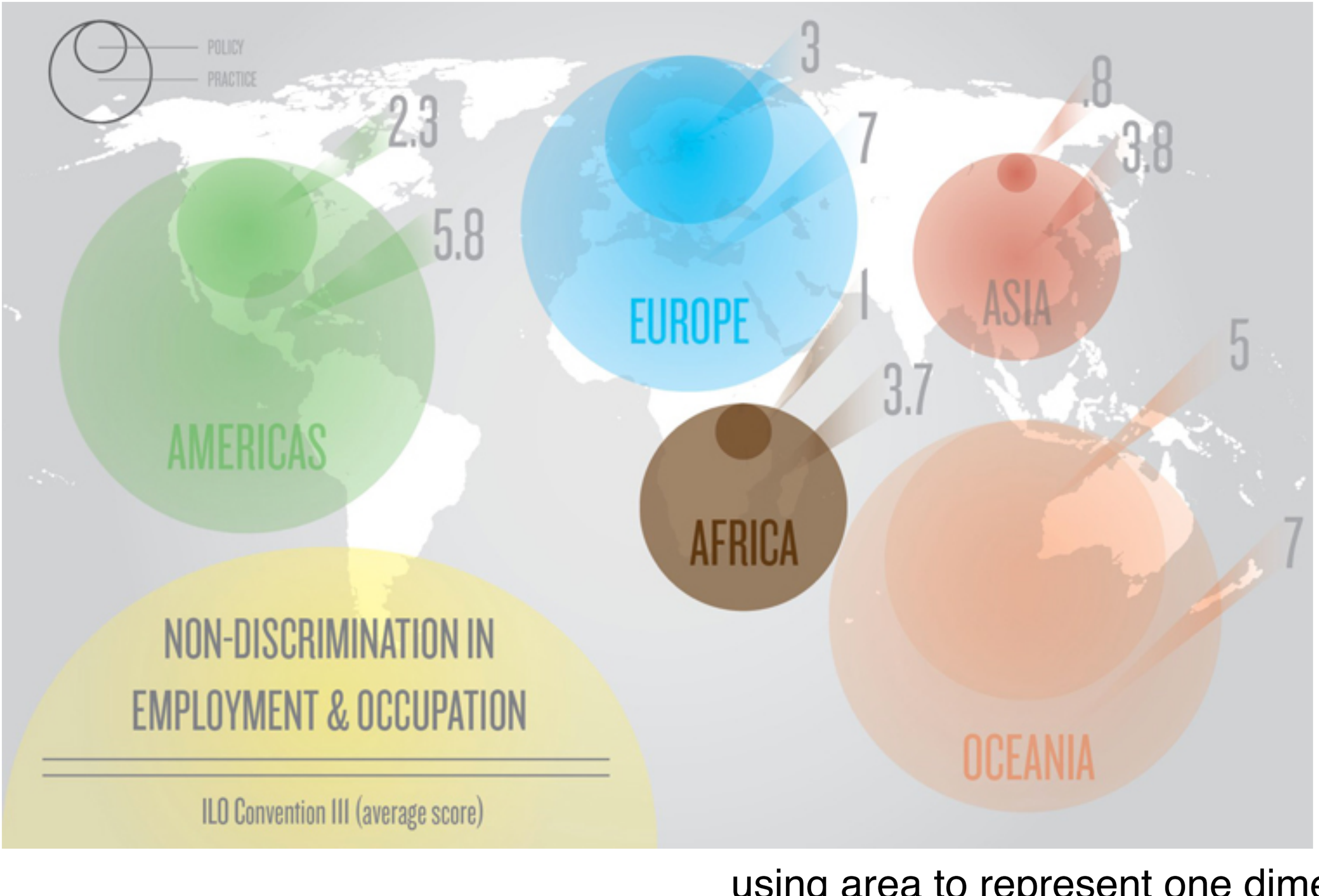


2. “Huge differences” fail





using area (2 dimensions)



using area to represent one dime

Quiz: How does this fail?

THE ISSUE OF TRUST

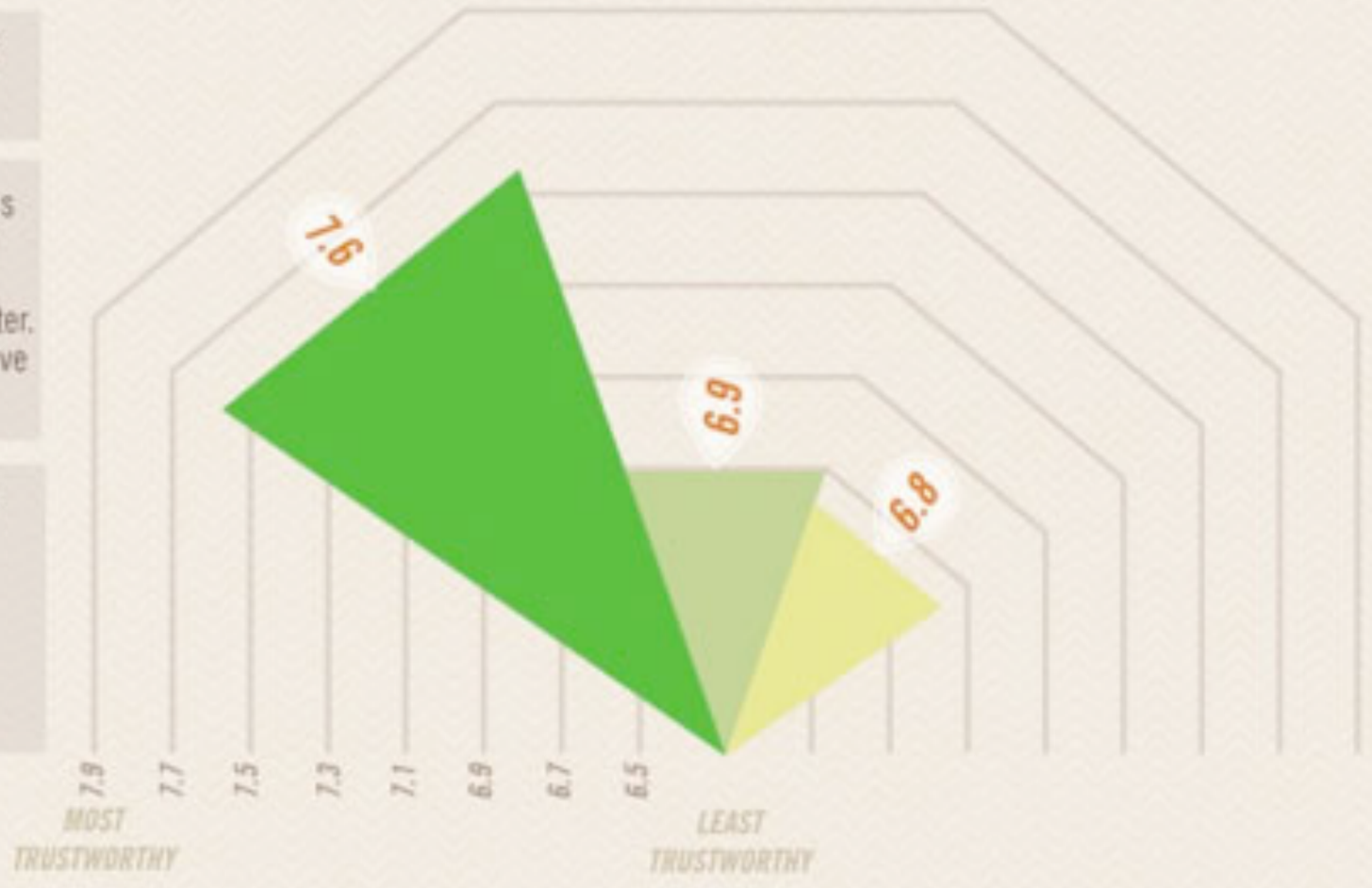
ACCENTS AND DISTRUST

Another reason why accents affects customer service is the question of credibility. If I can not understand you, then I can not trust you.

An experiment conducted by the University of Chicago demonstrated this aspect. The question posed, do trivia statements sound less true when spoken by a non-native speaker? Furthermore, listeners were told in advance that all of the trivia questions were provided by the experimenter. This way, even listeners who were knowingly prejudice against non-native accents should not have been affected.

The results showed that the heavier the accent the less trust worthy the person became.

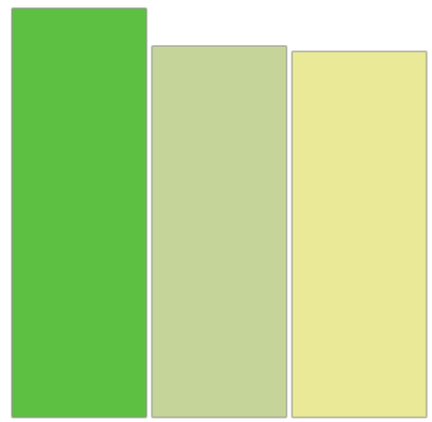
- NATIVE ACCENT
- MILD ACCENT
- HEAVY ACCENT



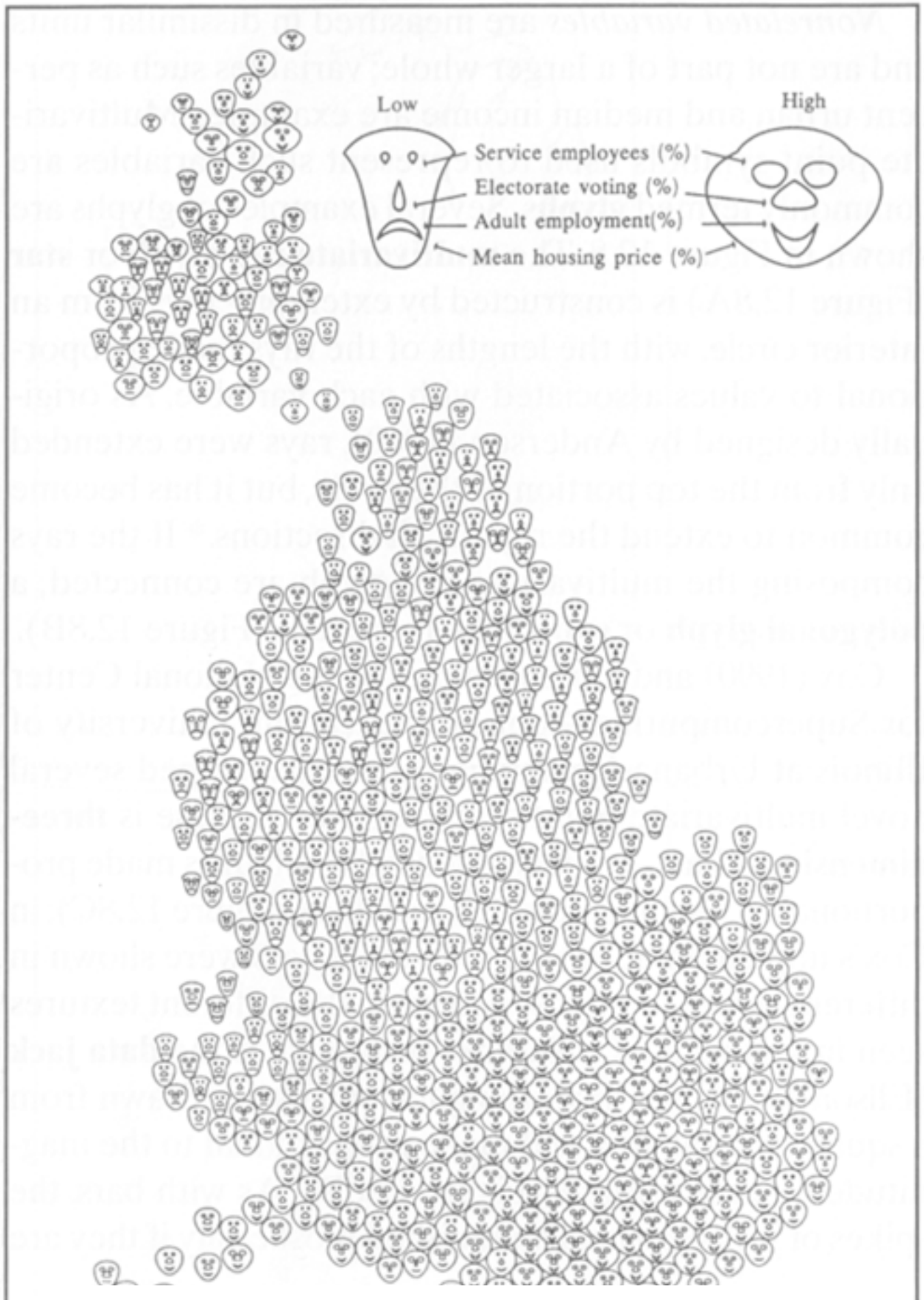
Distortion factor: 2.5

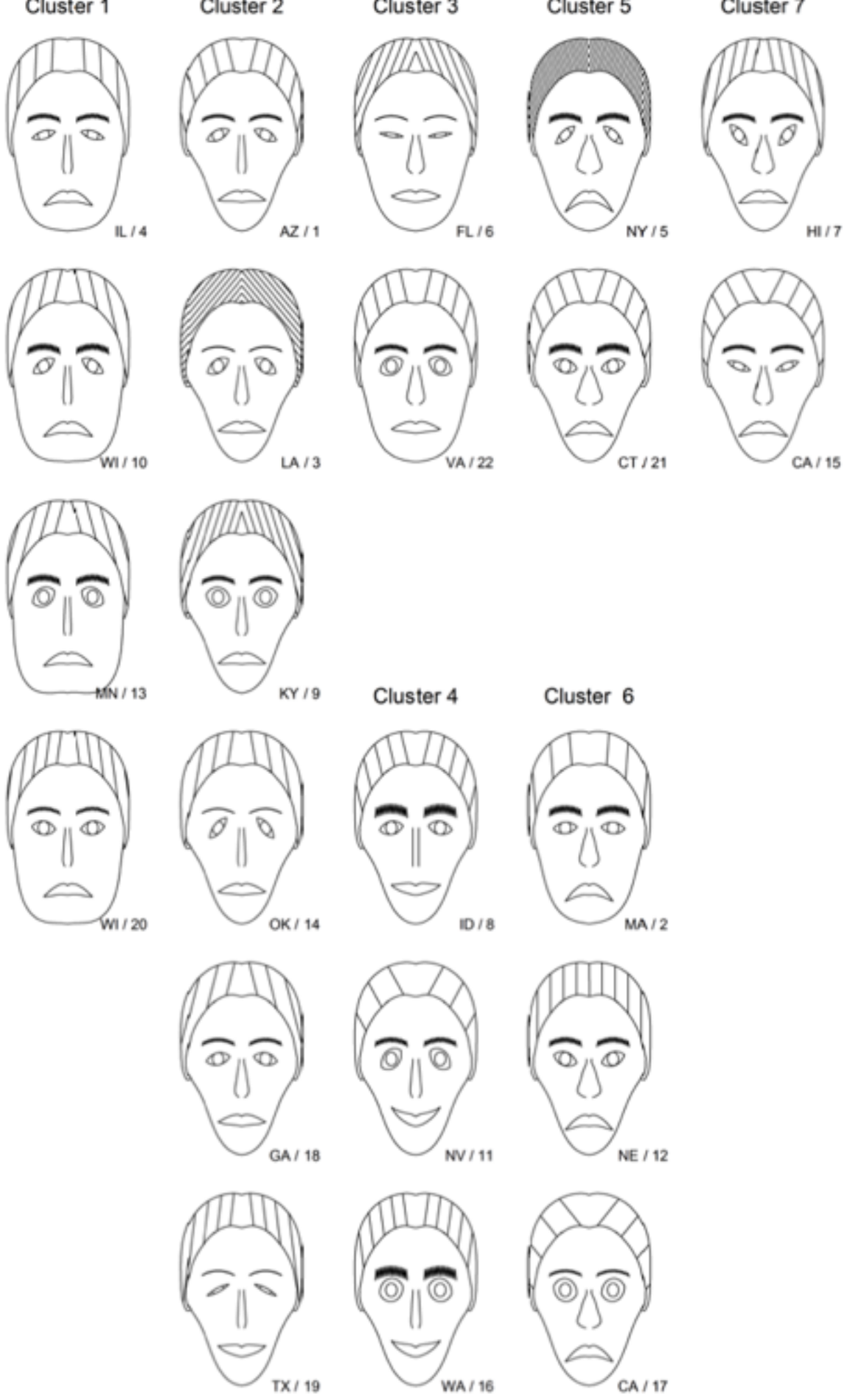
True data

Distortion factor: 5.0



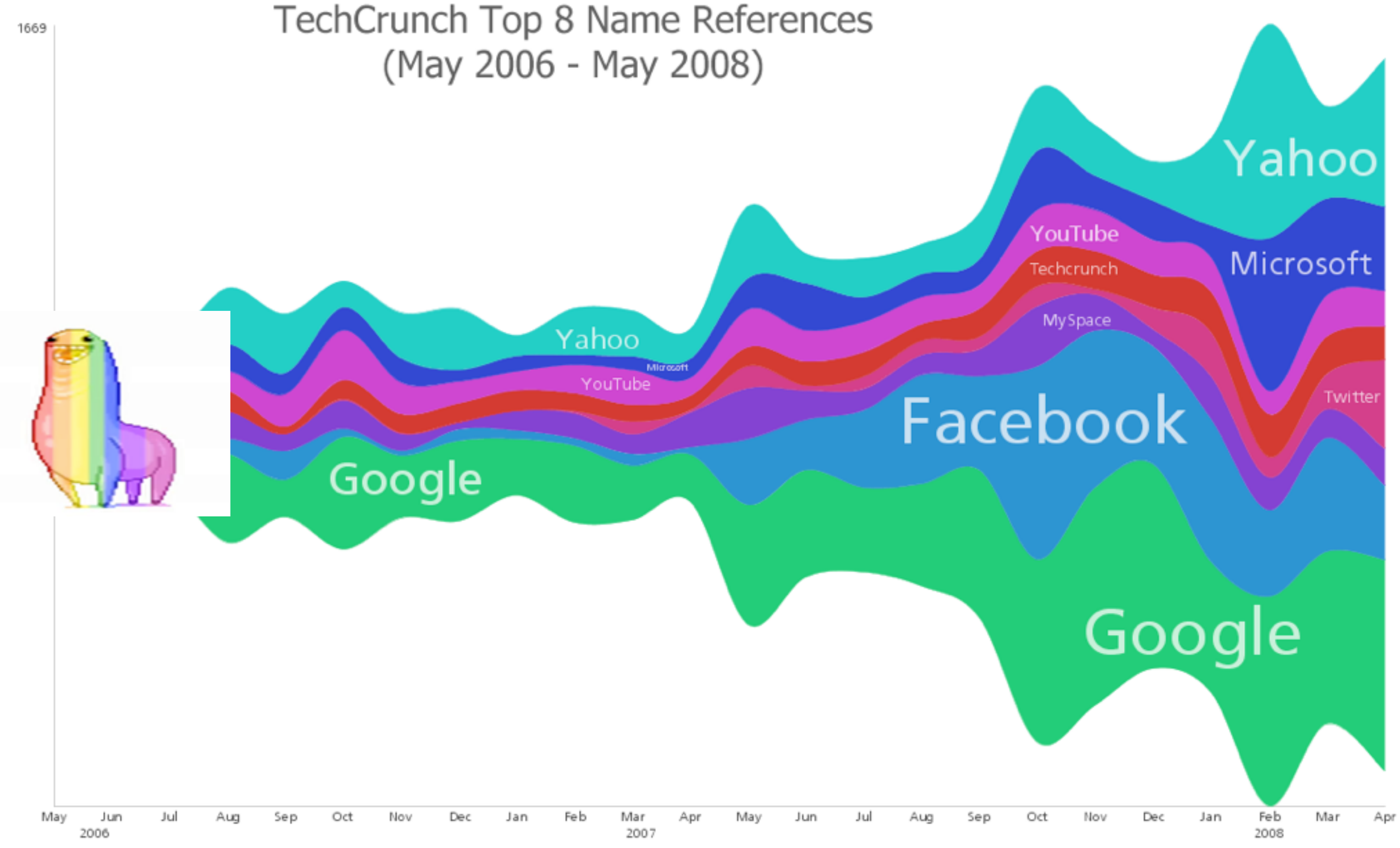
True data





Chernoff Faces

“streamgraphs” : double-stacked areas of horror



“abandon all hope ye who vieweth”