

part one!

biological basis of information design

introduction to

what visualisations can do for us

information

graphics

purpose - what is your (specific) goal?

& data

visualisation

data - what kind of data do you have?

visual dimensions - representing data visually

(@emax)

University of Oxford for
Open Data Institute Short Course

communication - deception and bad
infographics

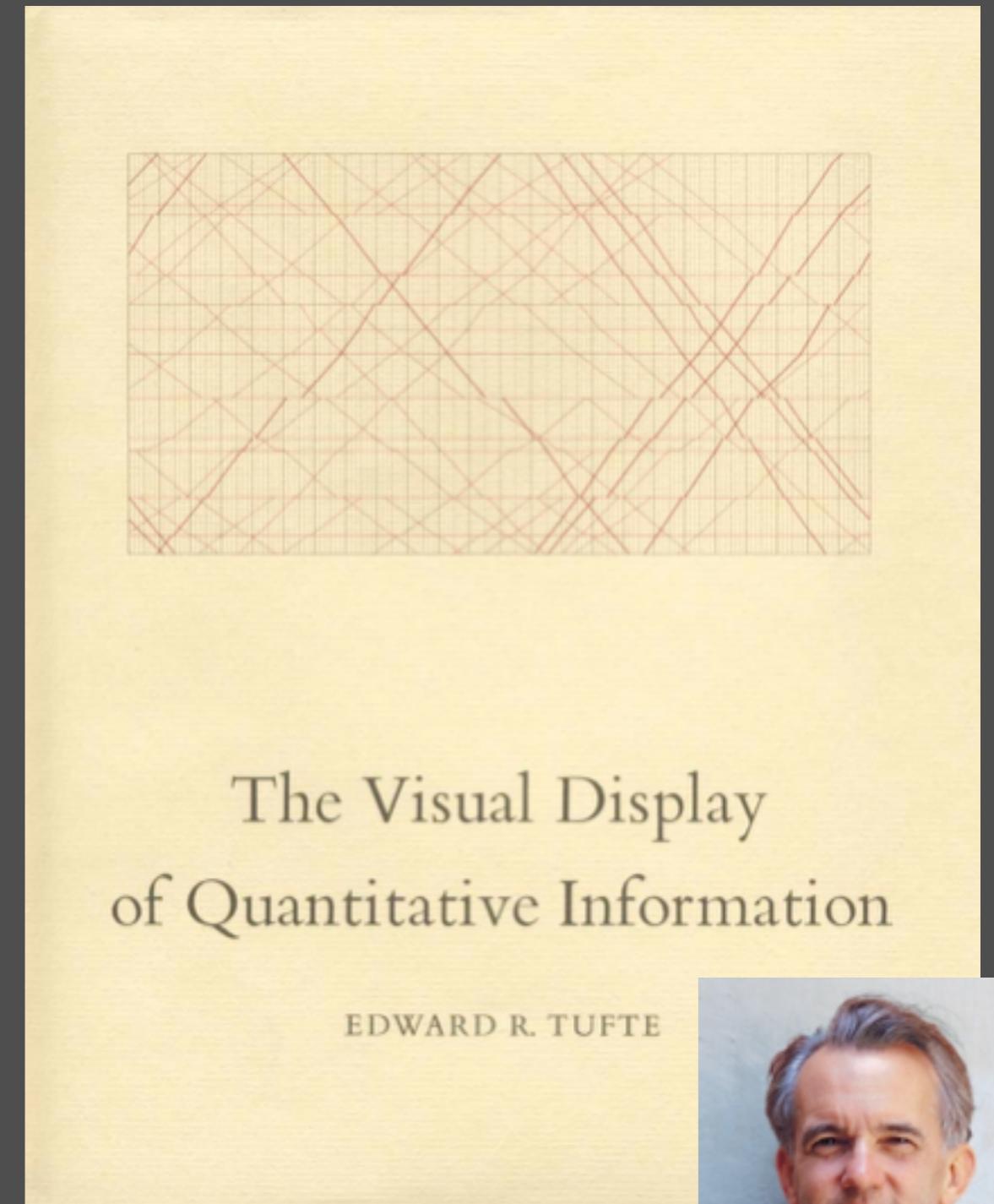
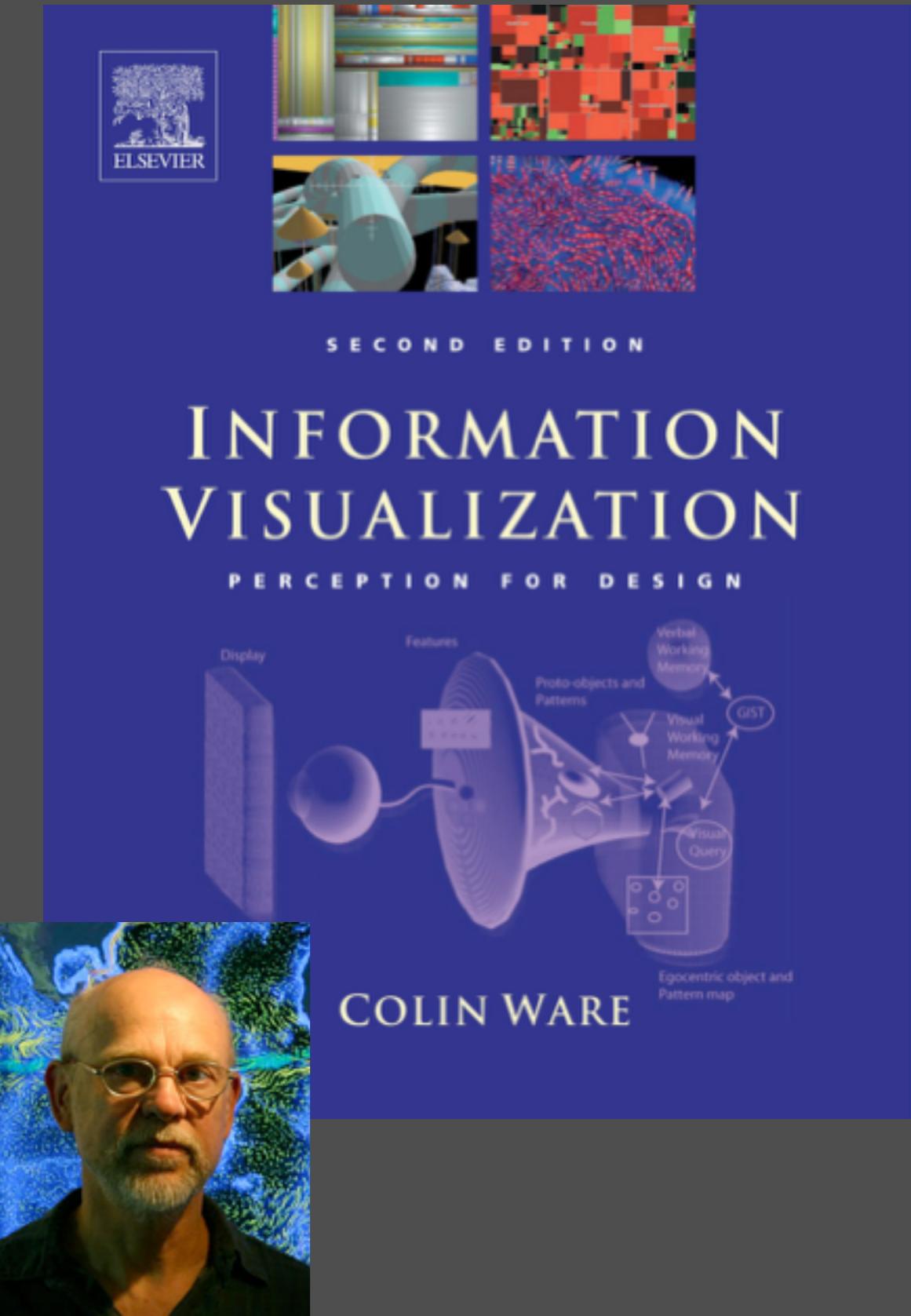
what are the **goals** of visualisation?

how do you **choose** a visual representation
for data?

how do you **evaluate** a visualisation?

key objectives

theory



praxis

ben shneiderman

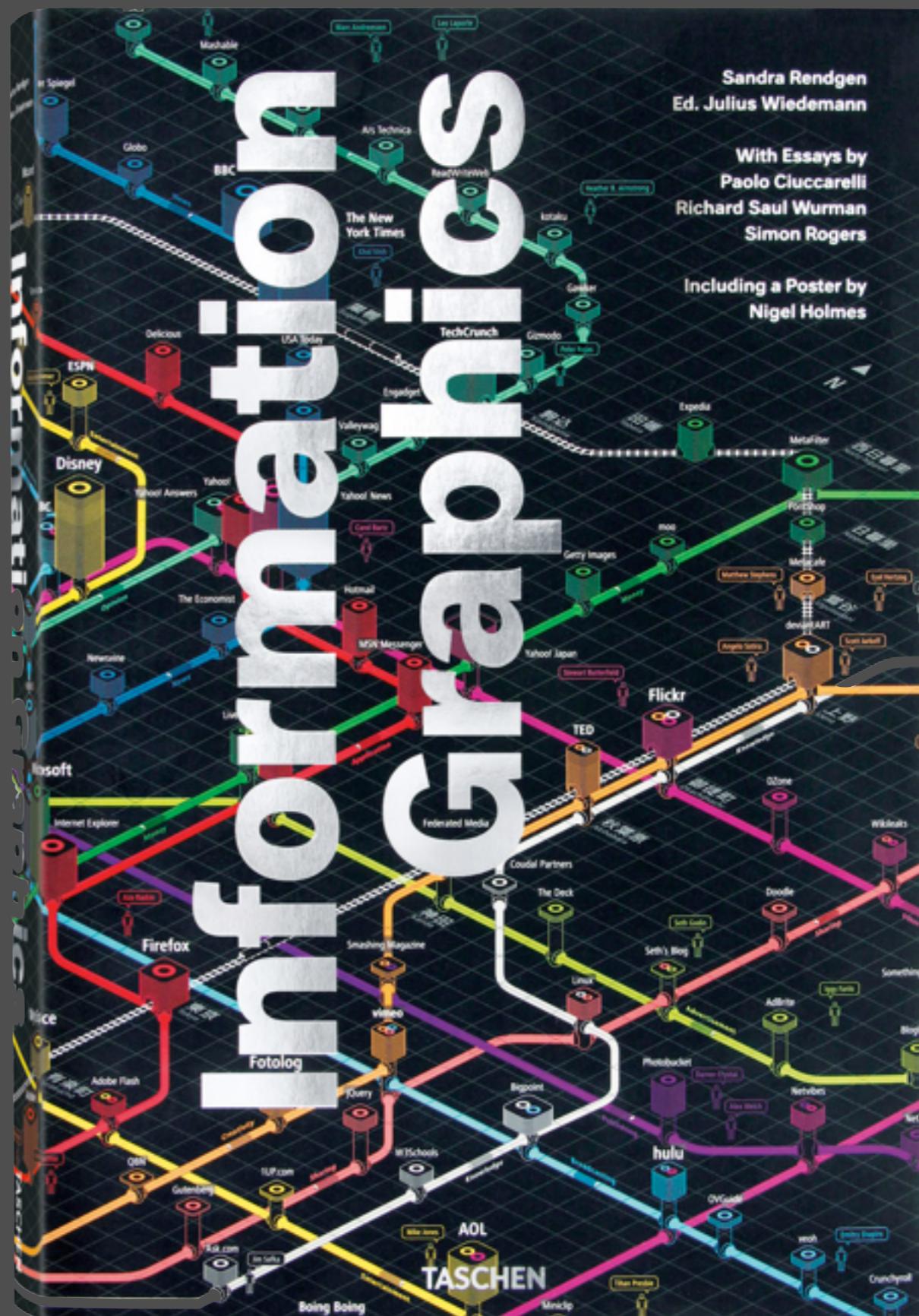
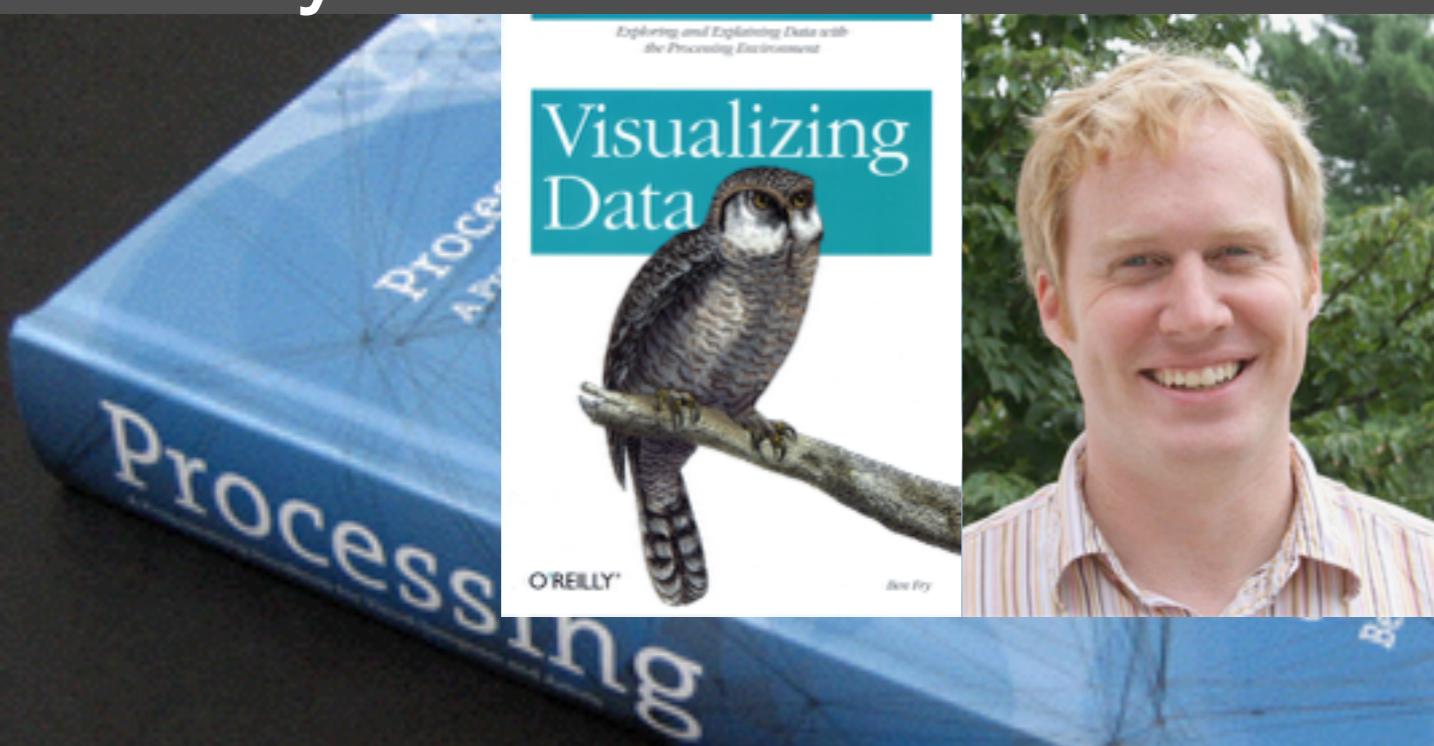
University of Maryland

READINGS IN
INFORMATION
VISUALIZATION
USING VISION
TO THINK

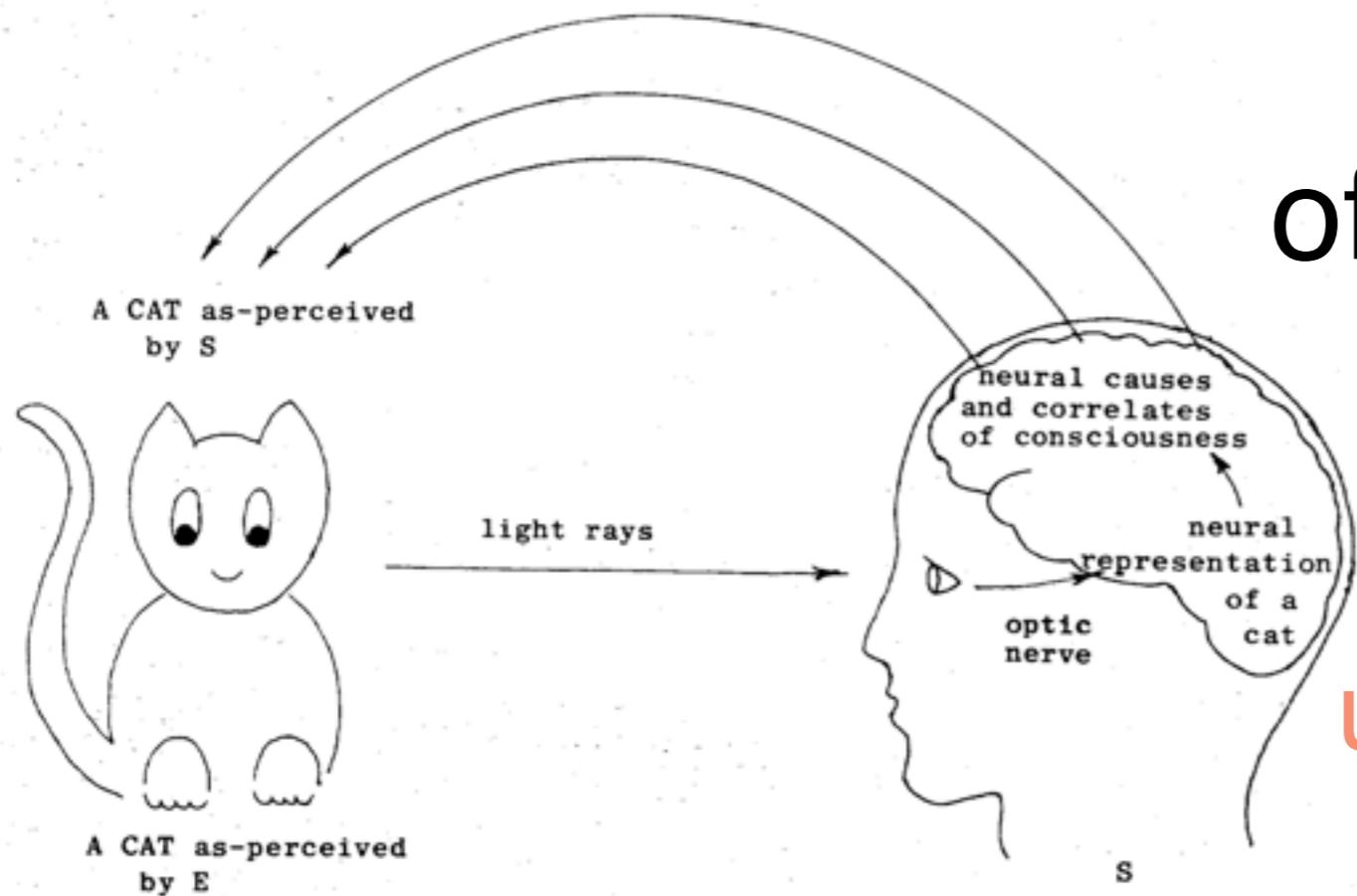


ben fry

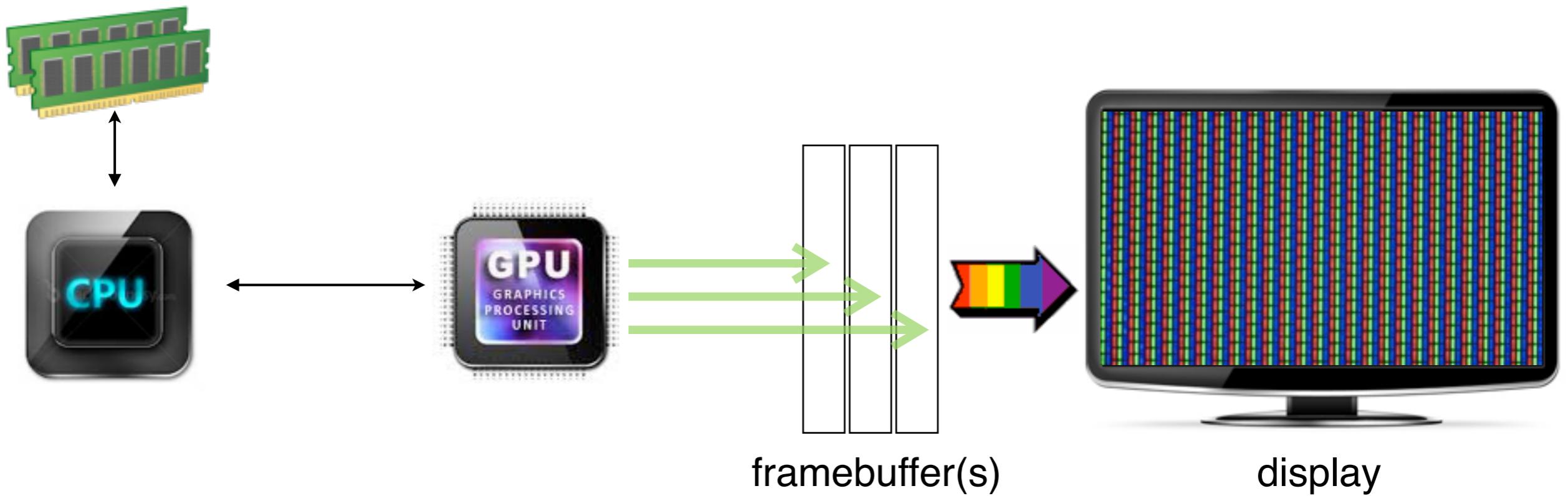
MIT Media Lab/fathom.info



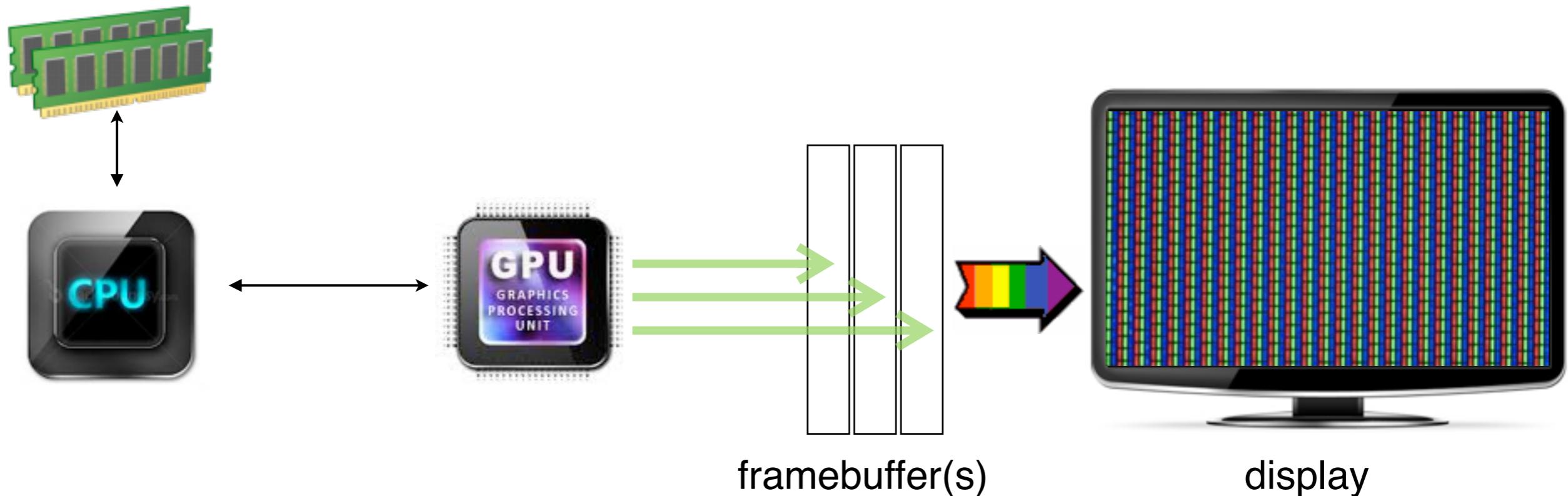
what is the goal of of information design?



1. to help people to understand & think about data.
2. to communicate facts



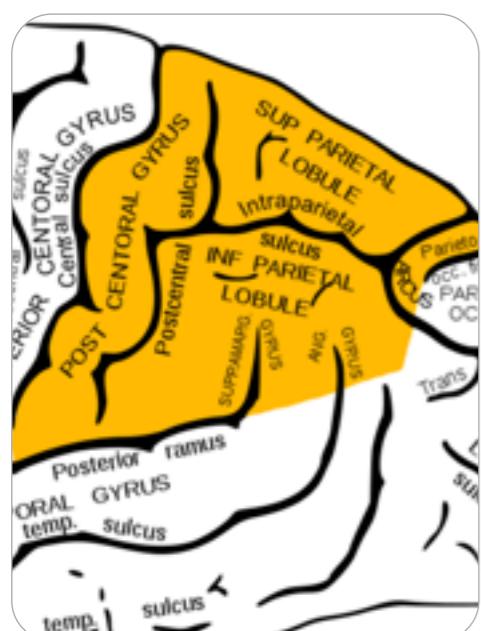
typical computer architecture



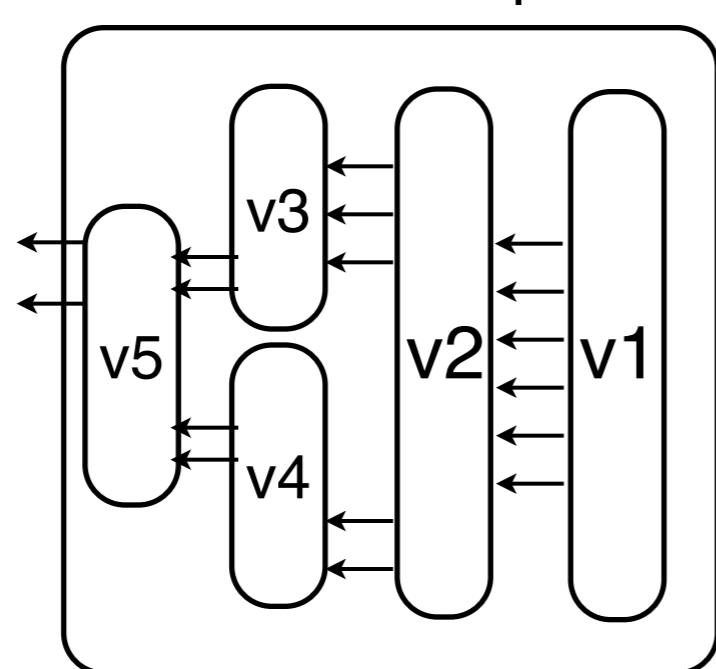
framebuffer(s)

display

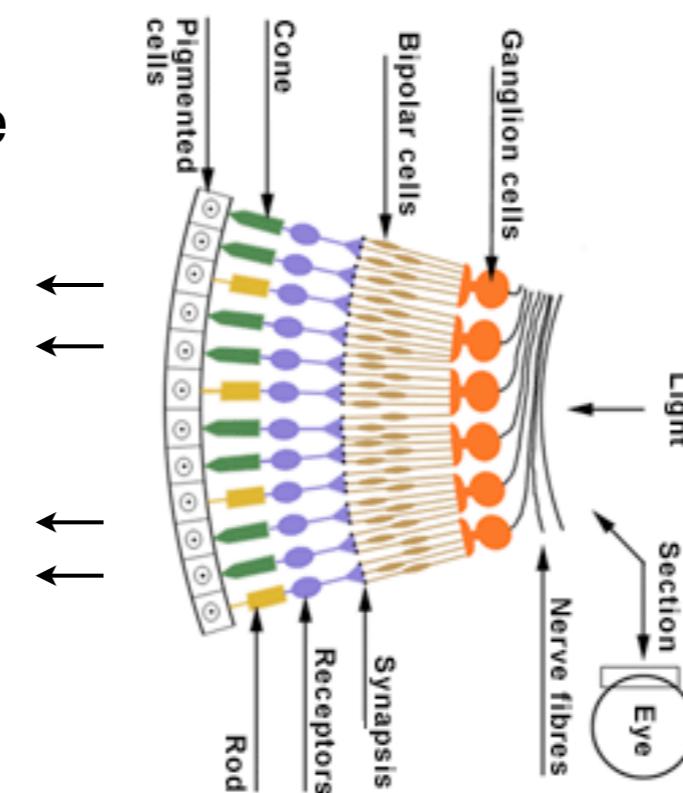
parietal lobe + frontal cortex



occipital lobe



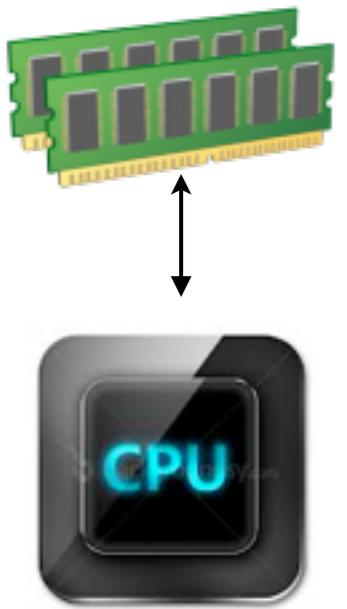
visual cortex
(pattern detection)



eye / iris / fovea

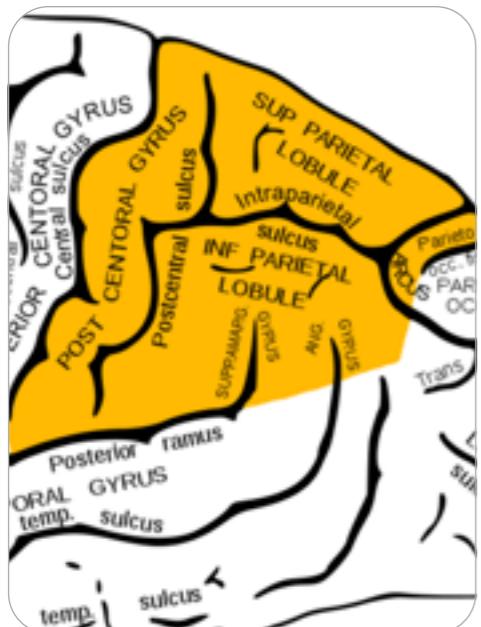
retina
(sensing)

spatial orientation
focus of attention
eye control,
perceptual fusion



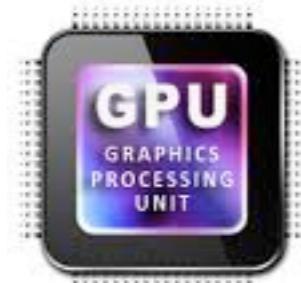
serial /
deliberative
processing
“attention-
focused”

parietal lobe + frontal cortex



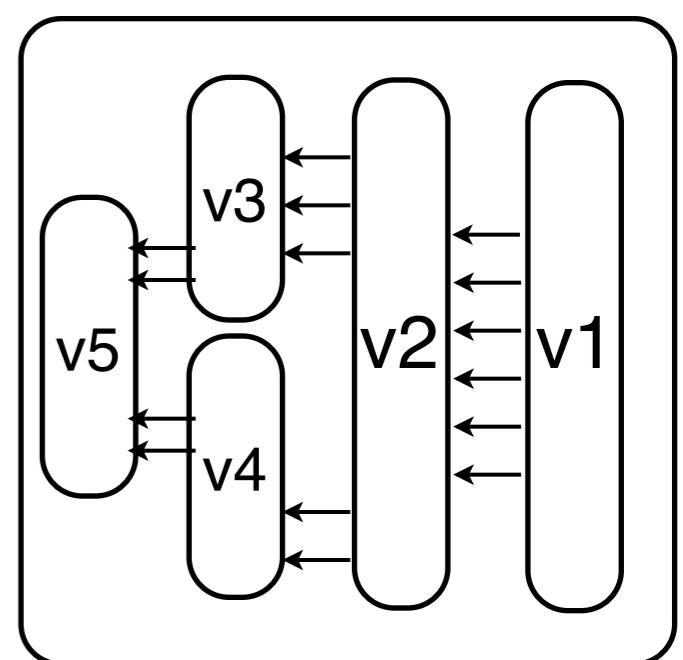
spatial orientation
focus of attention
eye control,
perceptual fusion

highly parallel

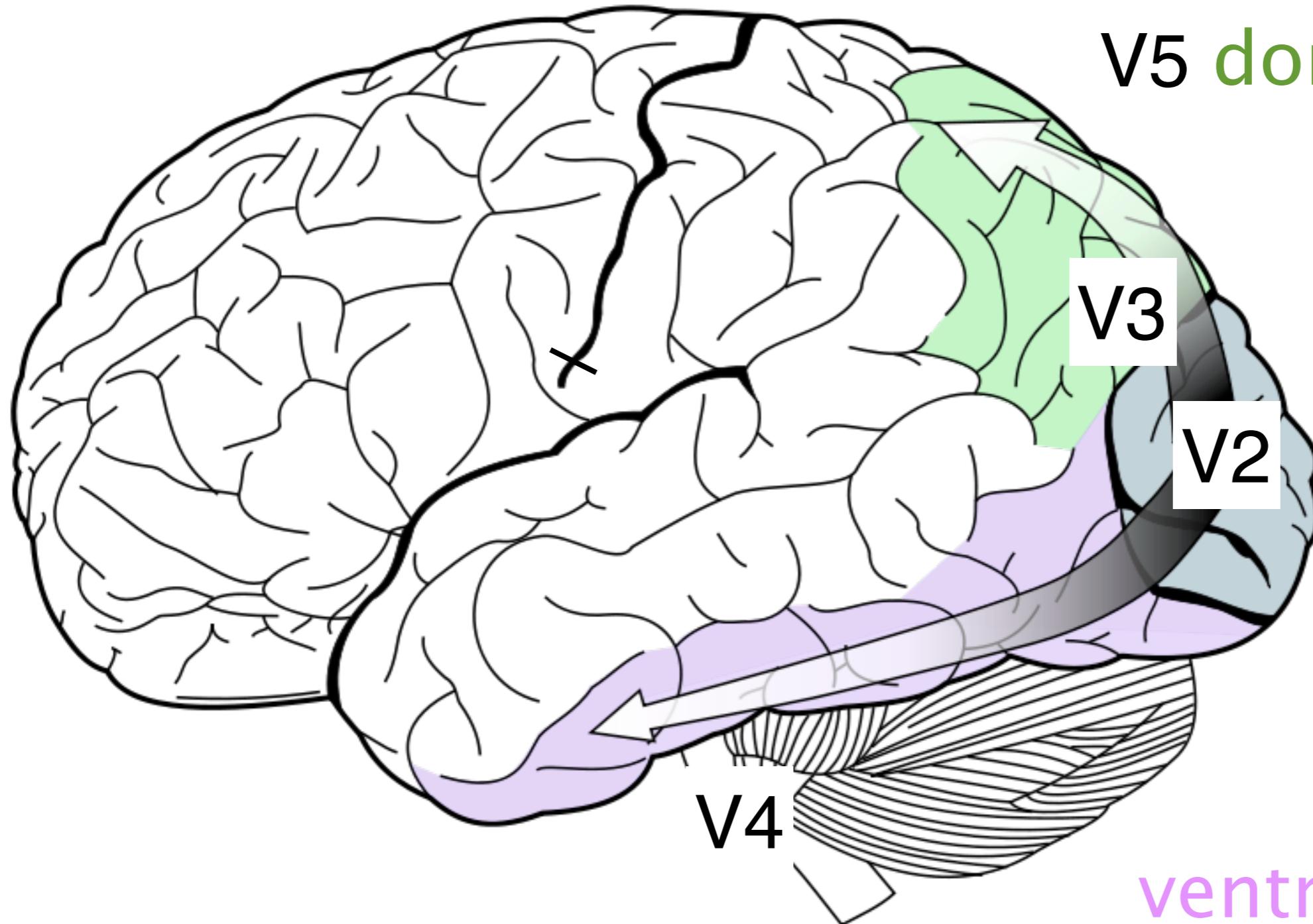


visual
processing
routines
optimised for
purpose

occipital lobe



visual cortex
(pattern detection)



V5 dorsal stream

"where/how"
pathway

V1
occipital
lobe

ventral stream

"what" pathway

object and person recognition

a

b

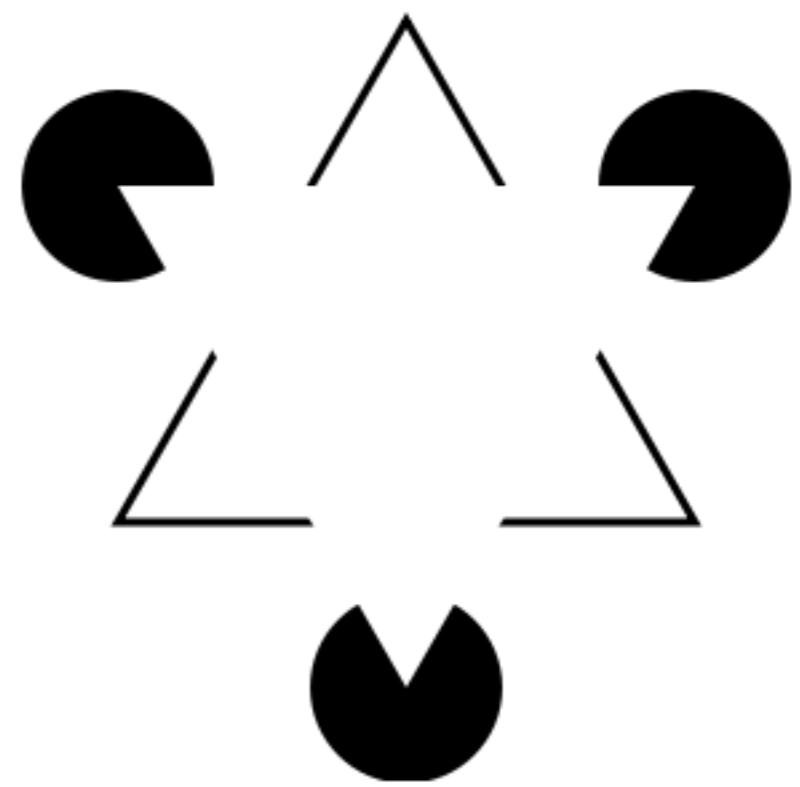
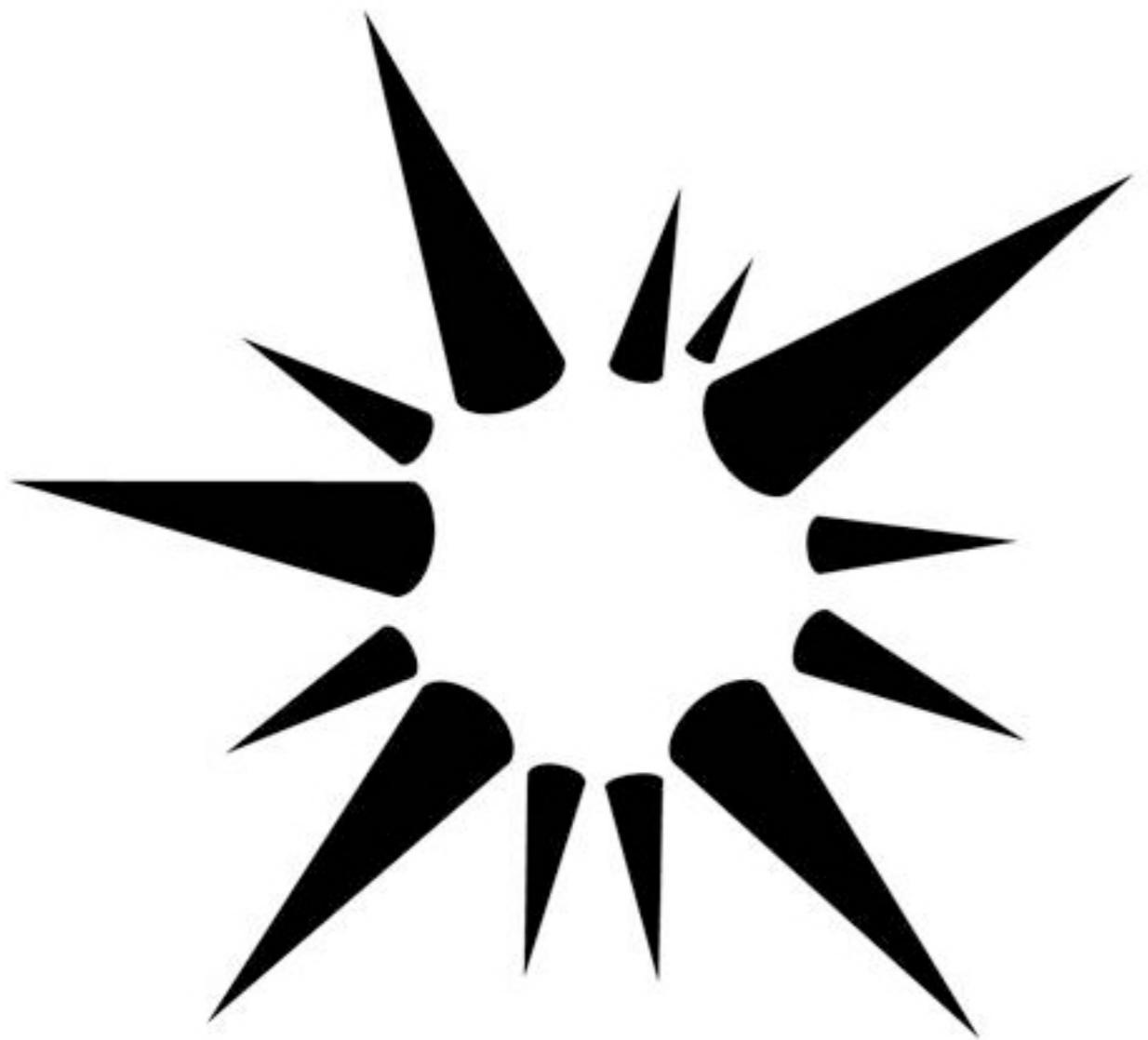
A large grid of 100 identical Persian characters 'ب' (Be) arranged in 10 rows and 10 columns. The characters are black and have a traditional Persian calligraphic style.

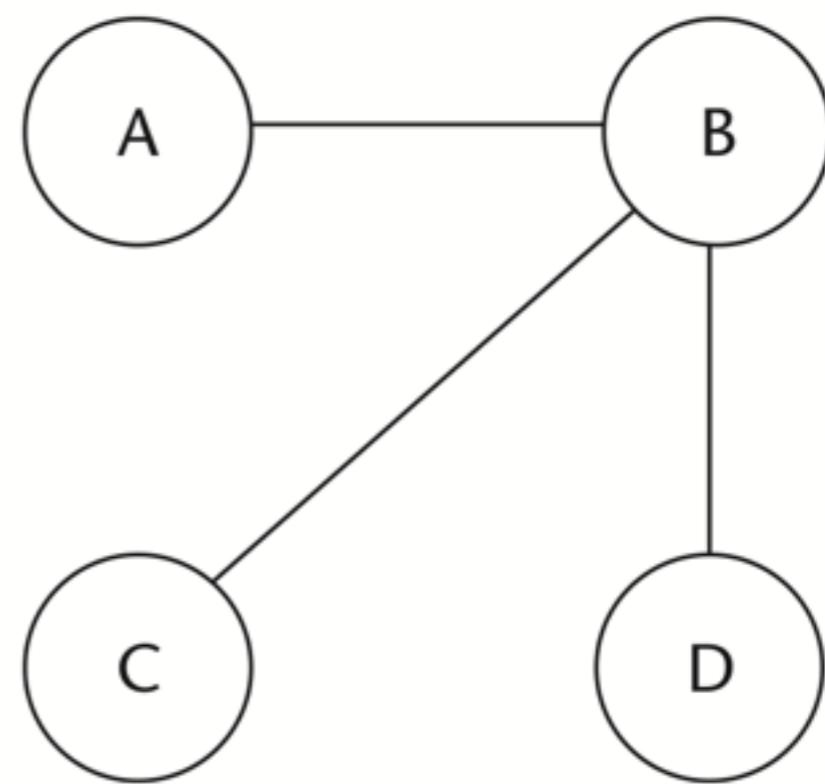
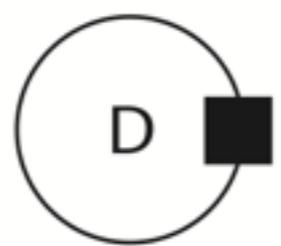
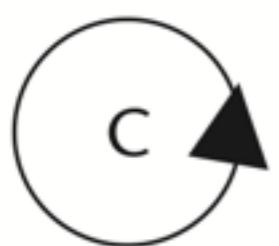
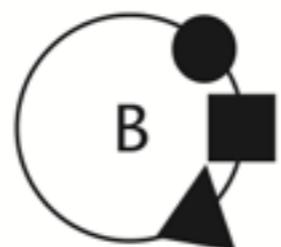
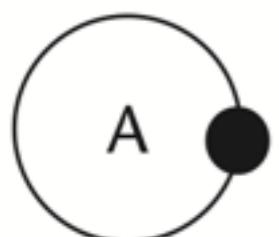
C

A large grid of handwritten cursive 's' characters, arranged in 10 rows and 10 columns. Each character is written in a distinct, flowing cursive style.

d

A large grid of 100 identical Y-shaped symbols arranged in 10 rows and 10 columns. Each symbol consists of a vertical line with a horizontal stroke extending from its right side. The symbols are evenly spaced and cover the entire area.





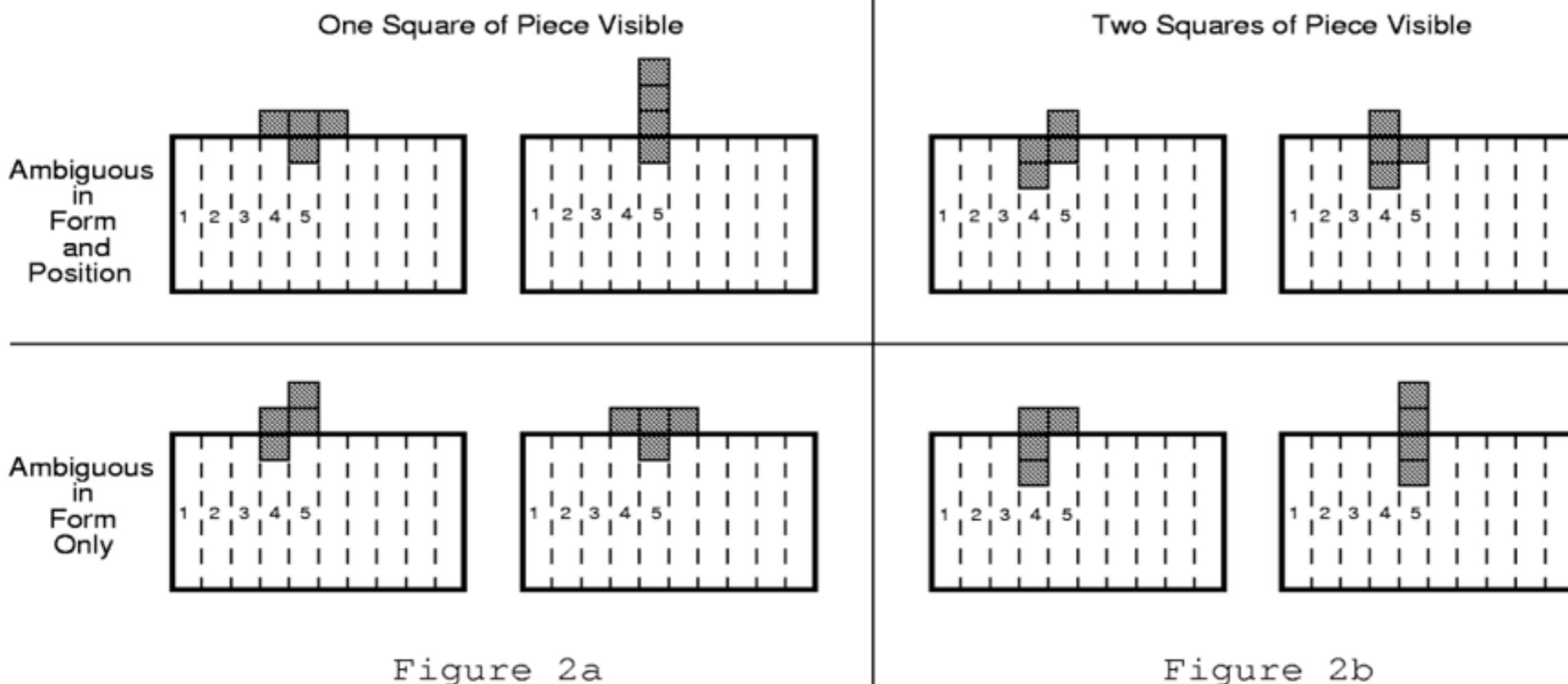
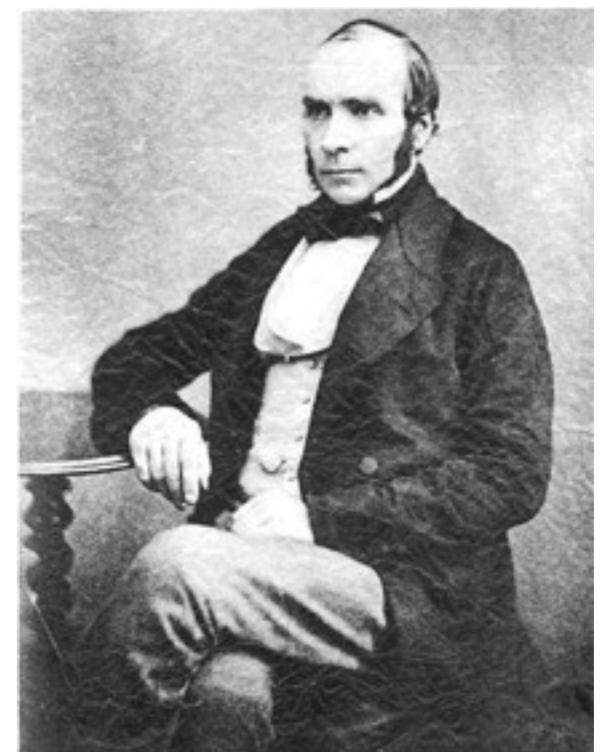


Figure 1. ILLUSORY RECOGNITION

50 0 50 100
X Pump • Deaths from cholera



London Cholera Outbreak John Snow, 1854

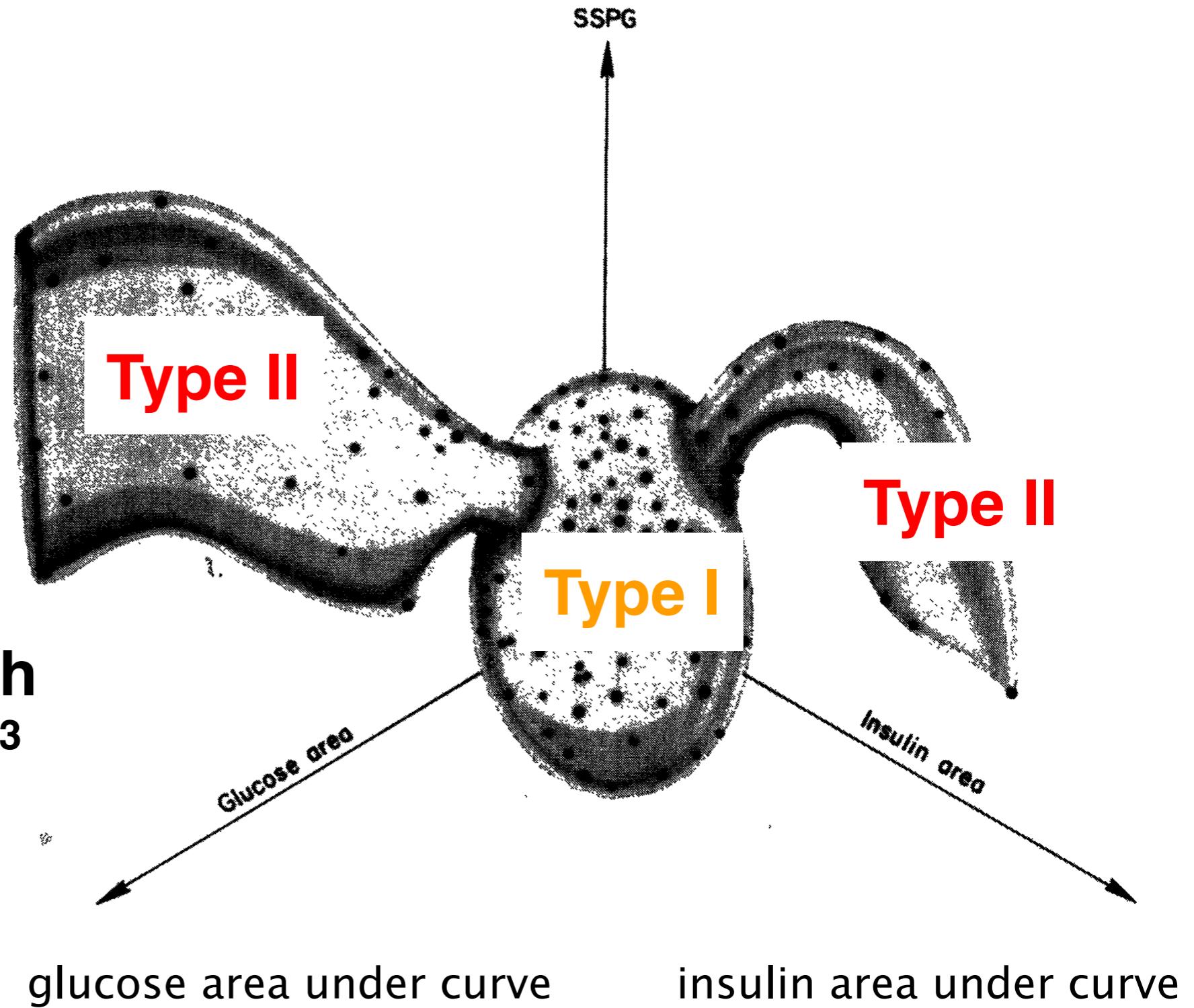


31 Aug 1854 - 127 deaths in 3
10 Sept - 500 deaths
End of outbreak - 616 deaths

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



steady state plasma glucose (response)

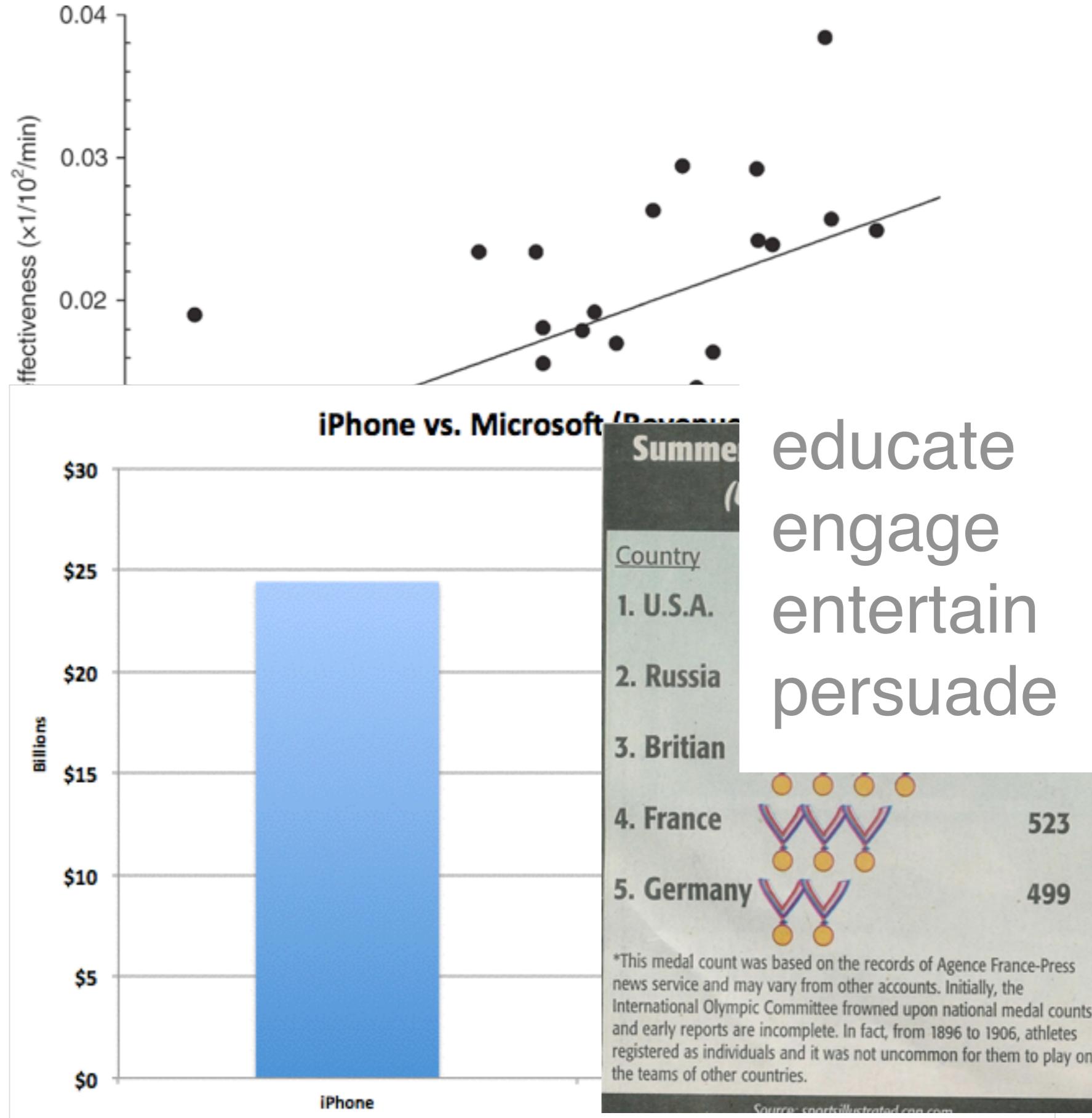


so how do we choose appropriate
visual representations
for our data?

1. purpose

understand

communicate



2. data types

$\{X_1, X_2, X_3, X_4, \dots\}$	X_i is...
$\{1, 200, 5, 6, \dots\}$	integral
$\{1.0, 2.0, 1.2, 4, \dots\}$	fixed point
$\{'a', 'b', '12C', 'd' \dots\}$	alpha(-numeric)
$\{20\%, 30\%, 1\%, 5\% \dots\}$	fractional
$\{ \text{pear}, \text{apple}, \text{kiwi}, \text{pineapple}, \dots \}$	categorical
$\{ f(\text{pear}, \text{apple}), g(\text{apple}, \text{kiwi}), q(\text{kiwi}, \text{pineapple}) \dots \}$	relational

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

2. data types

$\{X_1, X_2, X_3, X_4, \dots\}$ X_i is...

$\{1, 200, 5, 6, \dots\}$ integral

$\{1.0, 2.0, 1.2, 4, \dots\}$ fixed point

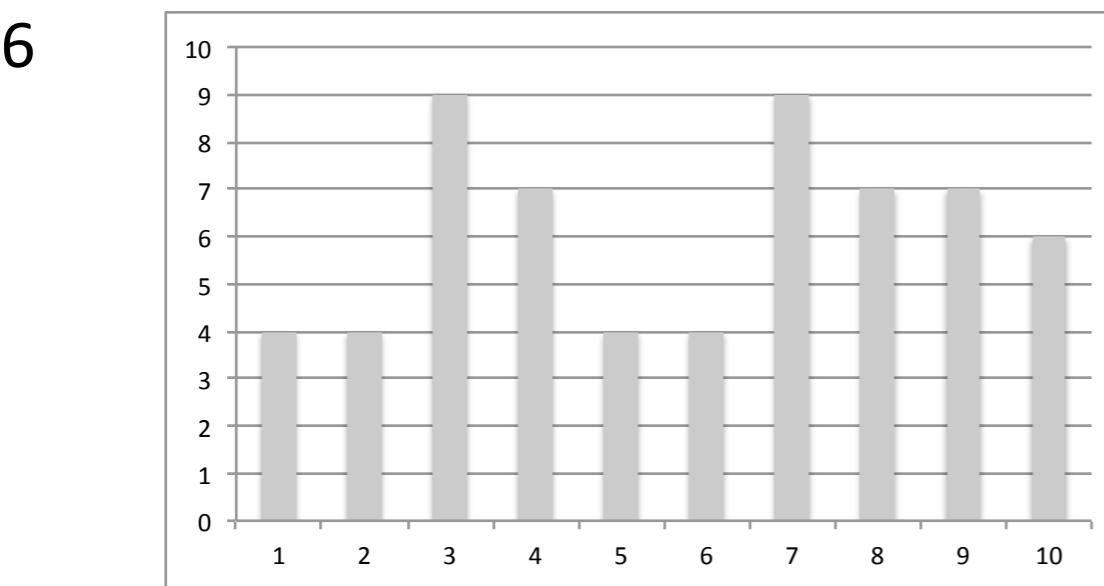
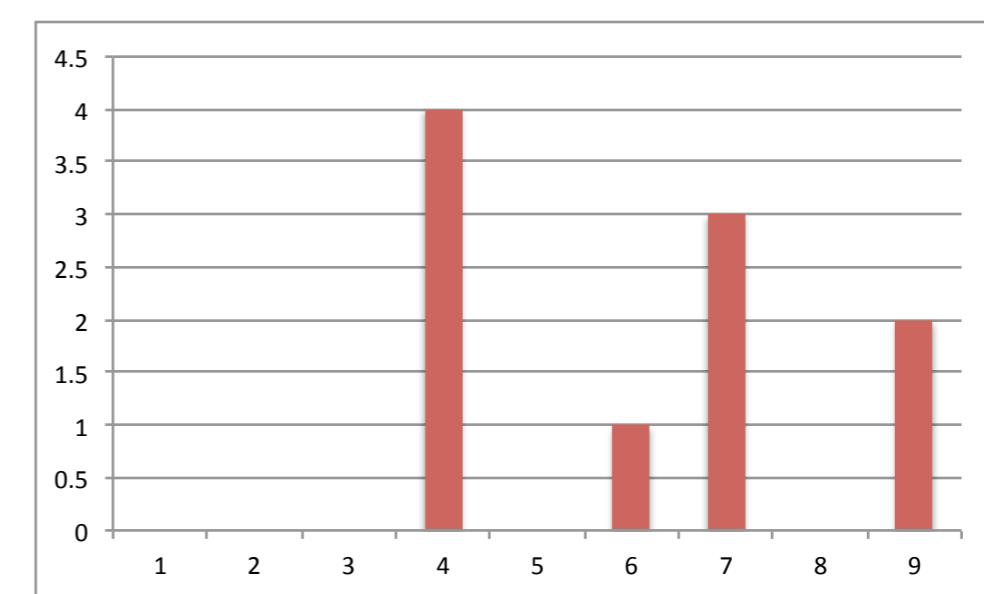
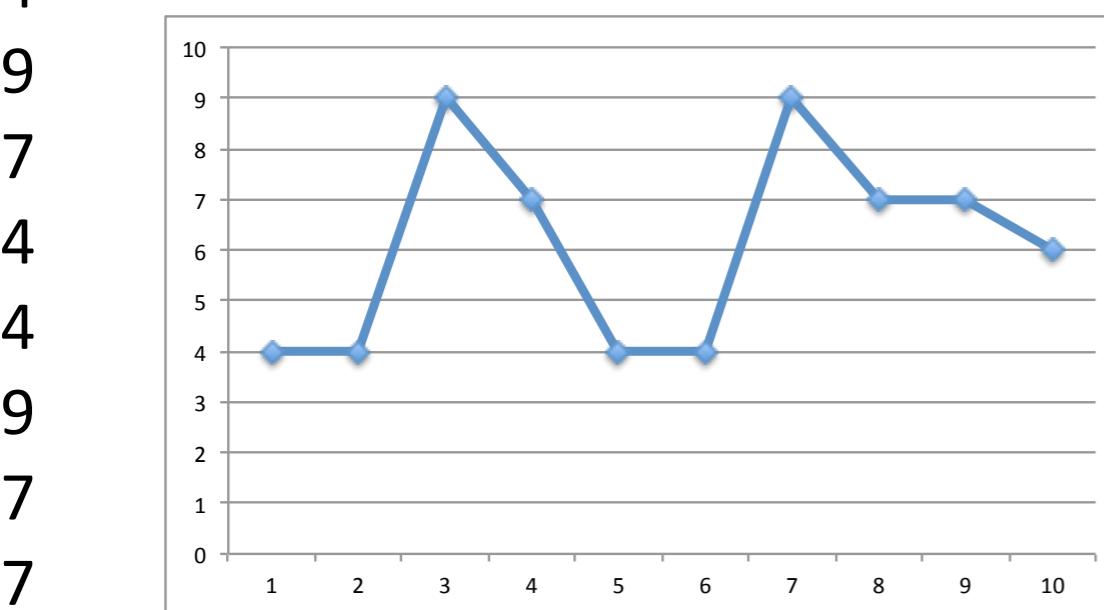
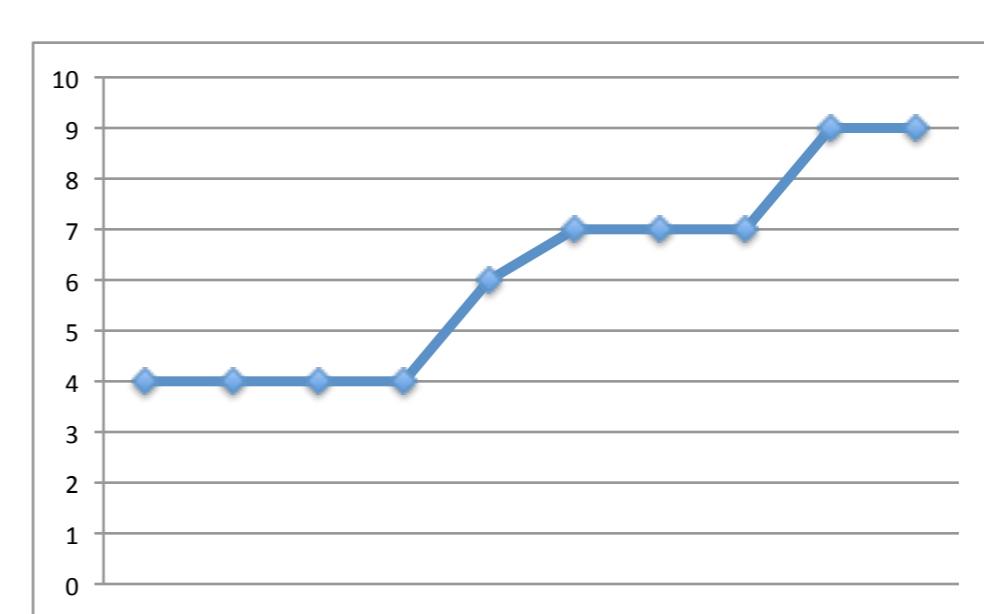
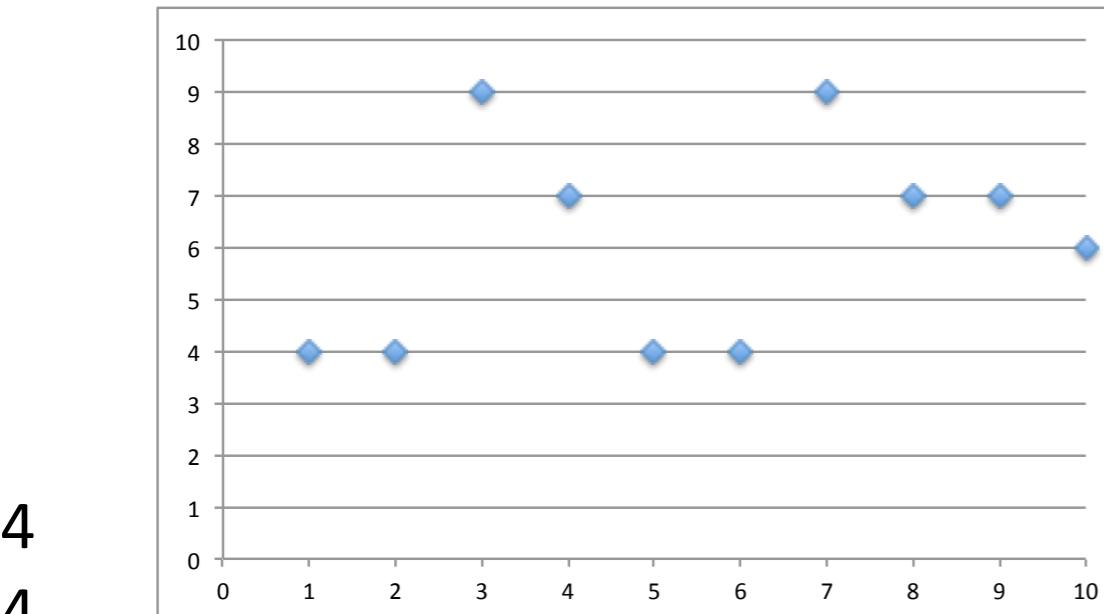
$\{'a', 'b', '12C', 'd' \dots\}$ alpha(-numeric)

$\{20\%, 30\%, 1\%, 5\% \dots\}$ fractional

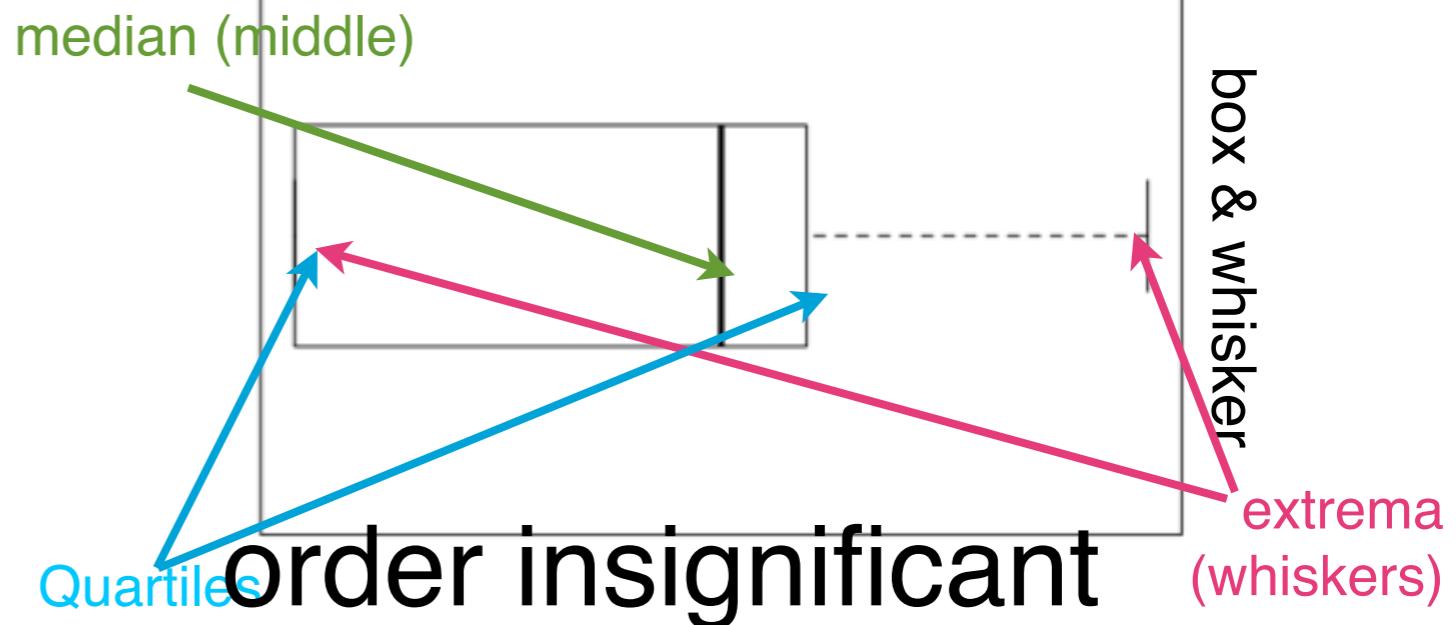
$\{\text{apple}, \text{orange}, \text{pear}, \text{pineapple}, \dots\}$ categorical

$\{f(\text{apple}, \text{orange}), g(\text{apple}, \text{orange}), q(\text{apple}, \text{orange}) \dots\}$ relational

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



ordering significant



sorted histogram

so you have a dataset...

it's probably multivariate

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

$$X = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ \text{pear}, & \text{kiwi}, & \text{pineapple} \\ t_1 & t_2 & t_3 \end{bmatrix} \dots$$

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"

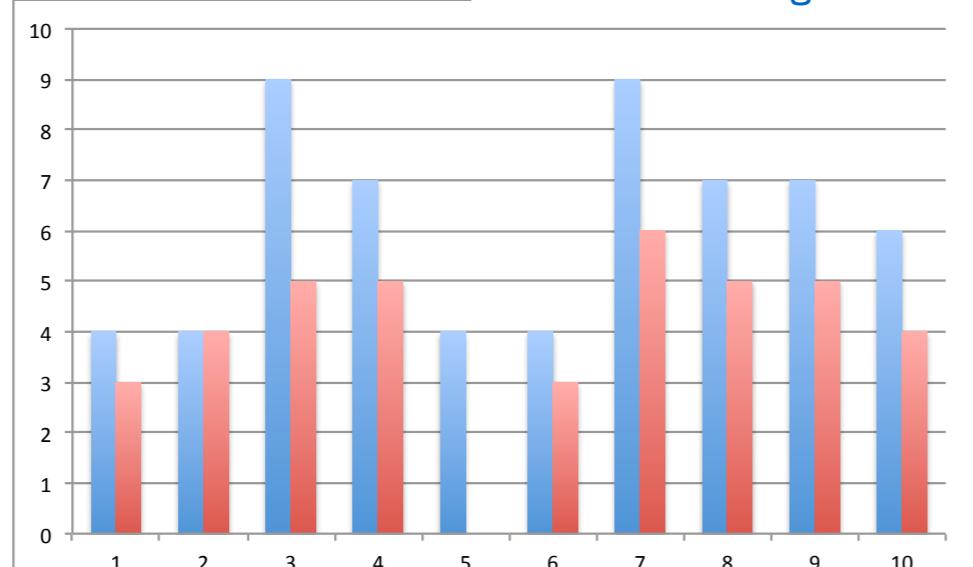
if these are observations of different things at a different points in time
"observations"

understanding objective(s) :

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

each dimension's variability

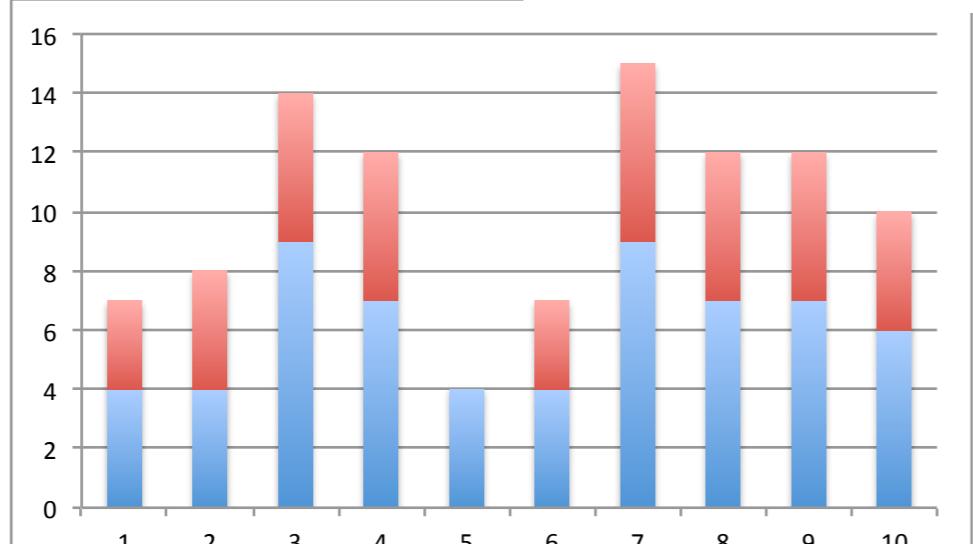
understanding elements



clustered bar

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

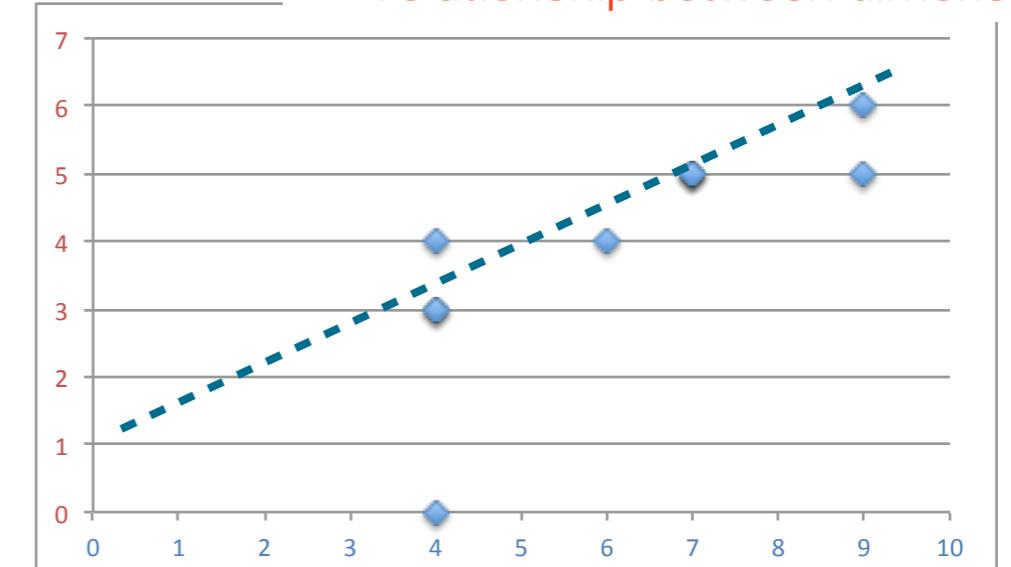
elements & their totals



stacked bar

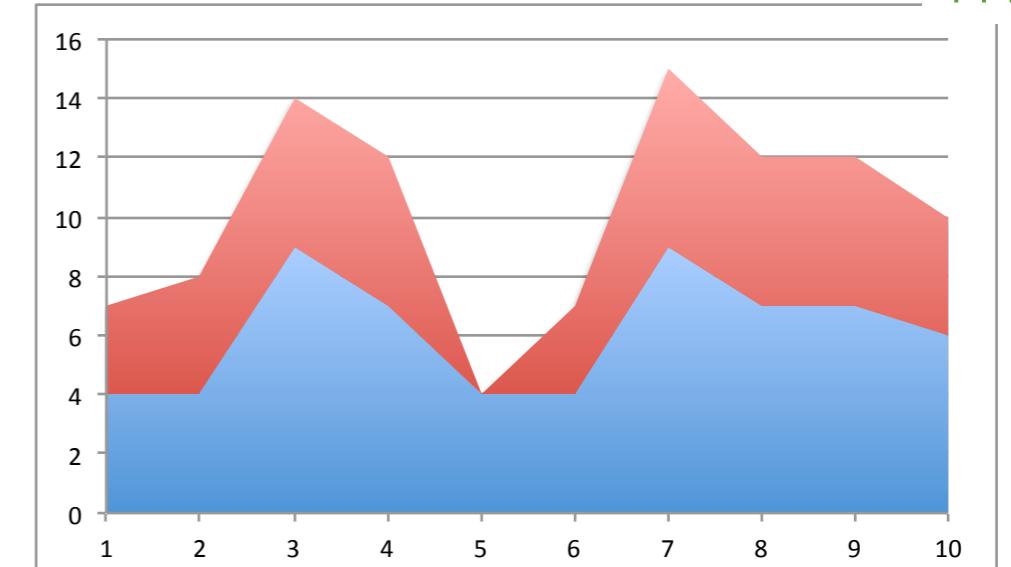
lines

relationship between dimensions

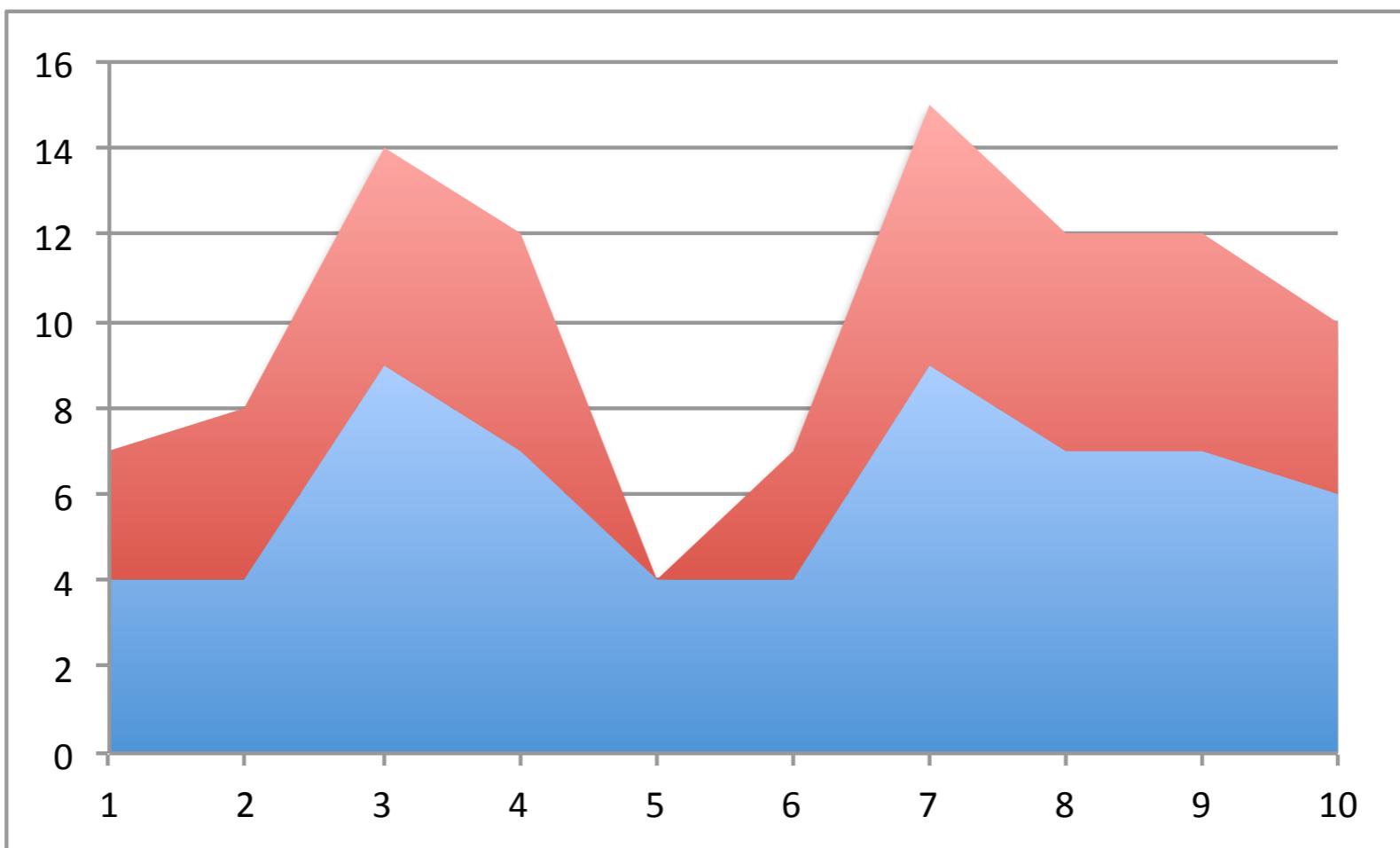
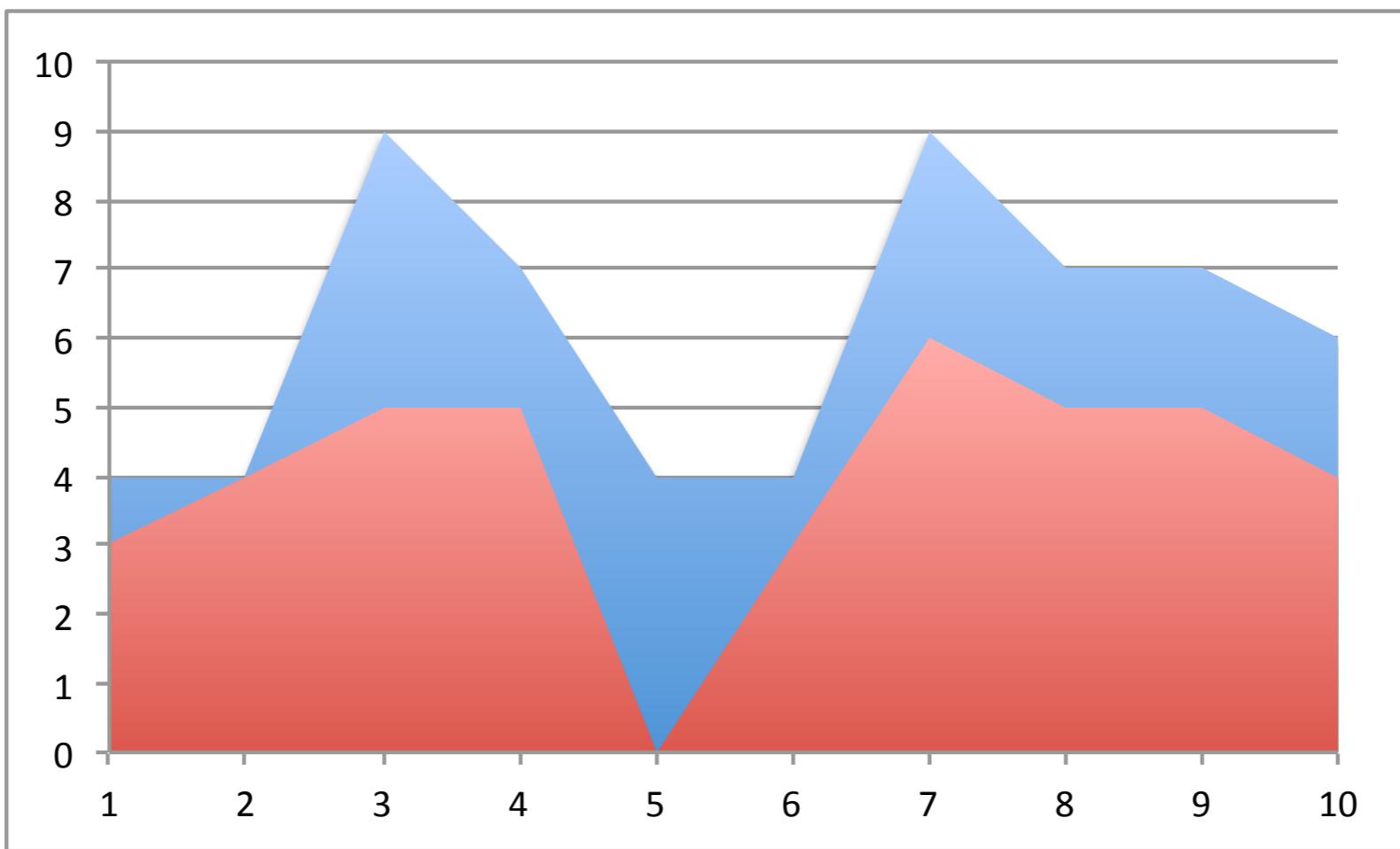


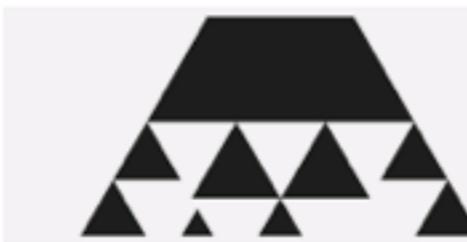
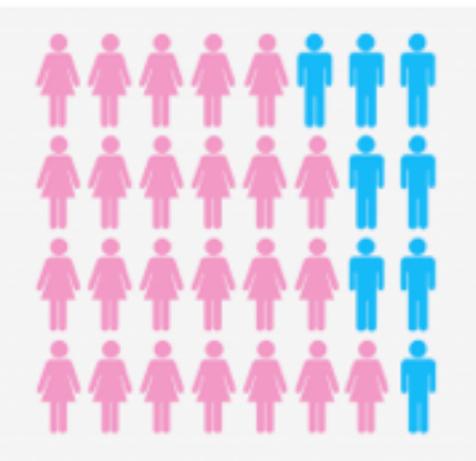
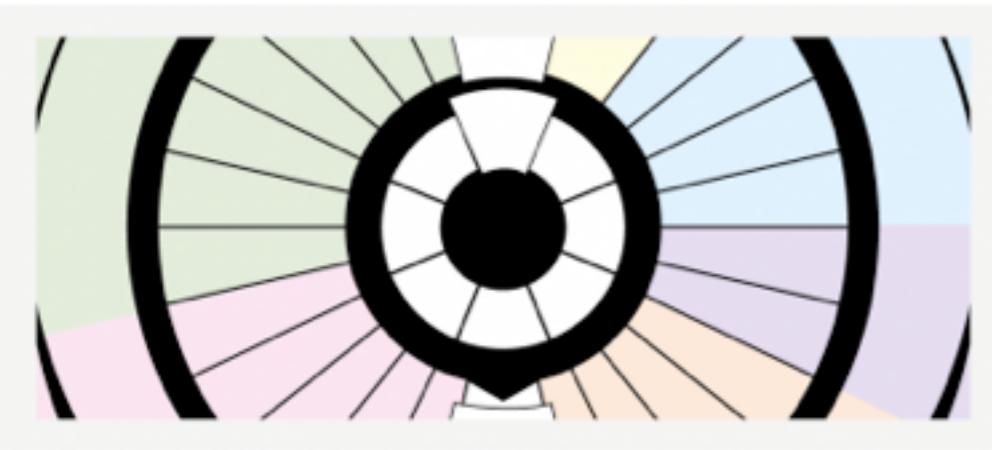
scatter

???



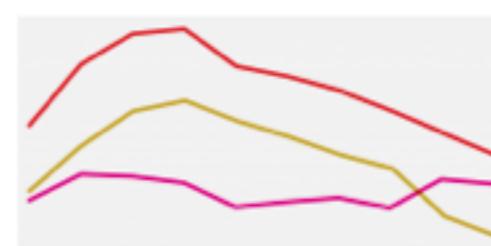
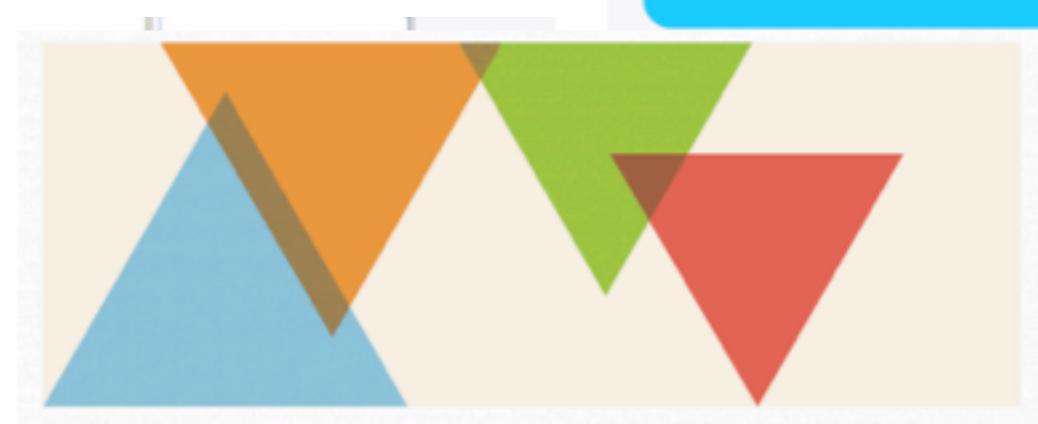
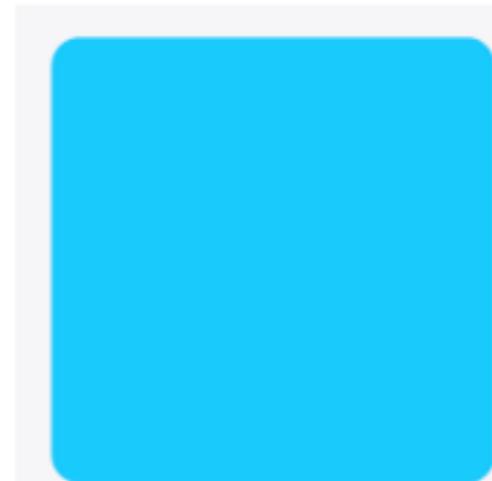
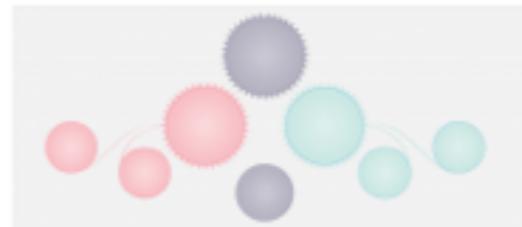
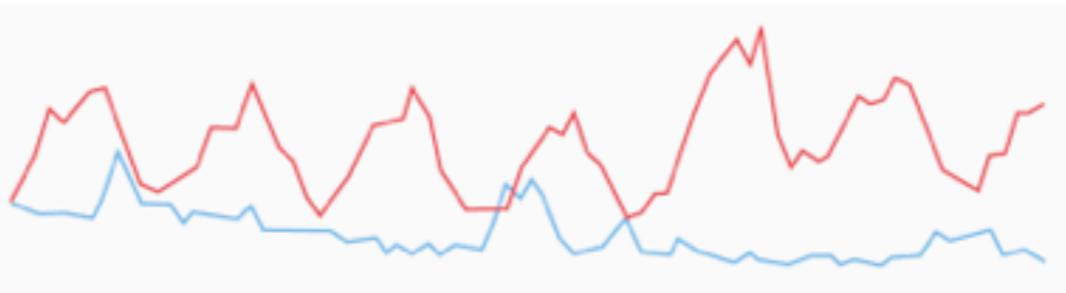
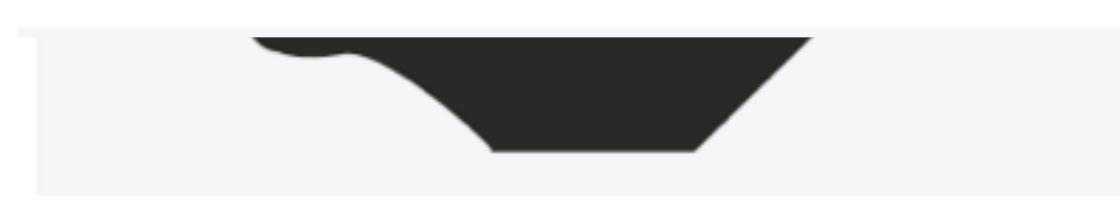
stacked area



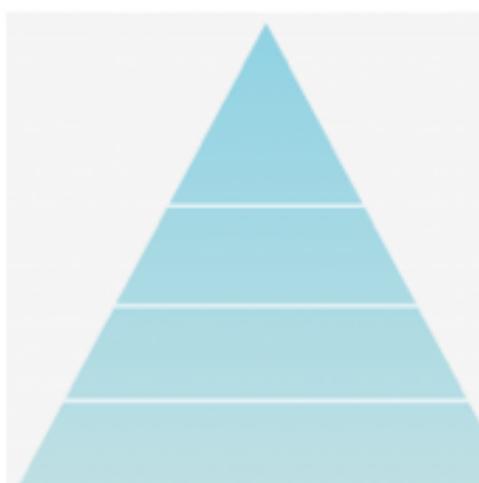
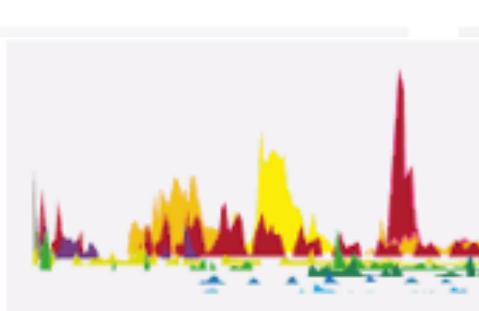


The Stranger Remembrance of Things Past
 Lord of the Flies One Hundred Years
 His Dark Materials
 The Handmaid's Tale
 The Sound and the Fury
 The Name of the Rose
 The Adventures of Huckleberry Finn
 The Rings To Kill a Mockingbird

3. Visual Dimensions



rible
 RAZZIE fantas
 COMPETI
 STRANG
 weird ODD
 ap half-baked
 IMPOSSIBL



data dimension types

integral

fixed point

alpha(-numeric)

ordinals

categorical

relational

...

visual dimension type

position

relative location
centrality



shape



colour

saturation
opacity

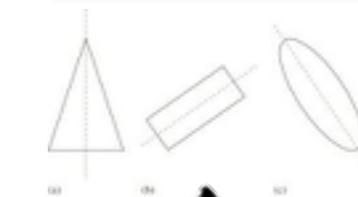


size

width
height



orientation



stroke

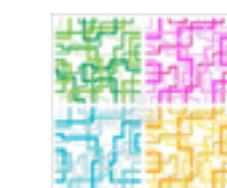
colour
pattern,
thickness



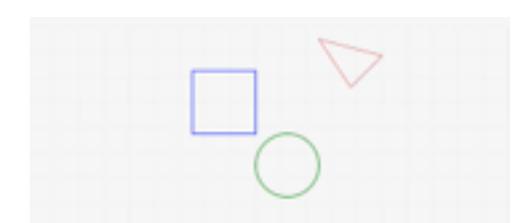
opacity



texture



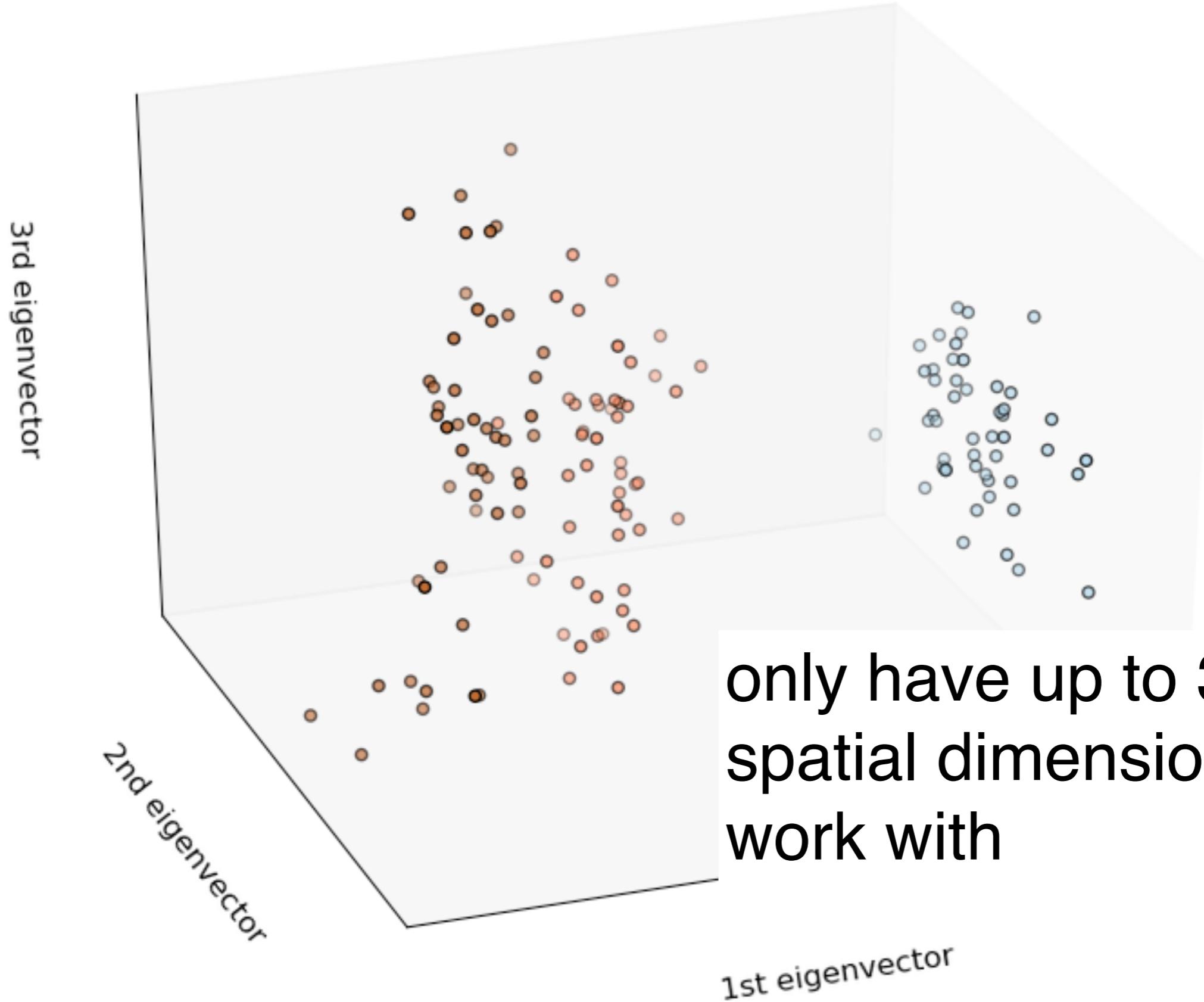
movement



juxtaposition

position

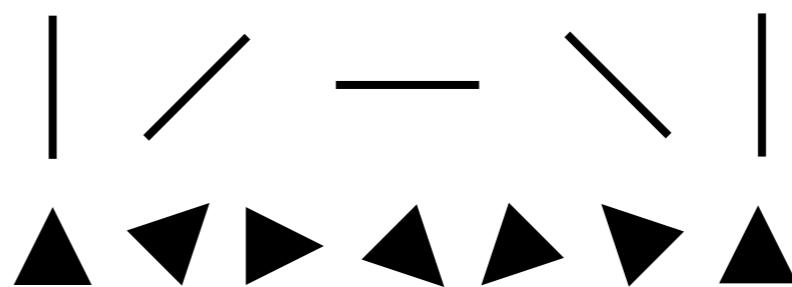
First three PCA directions



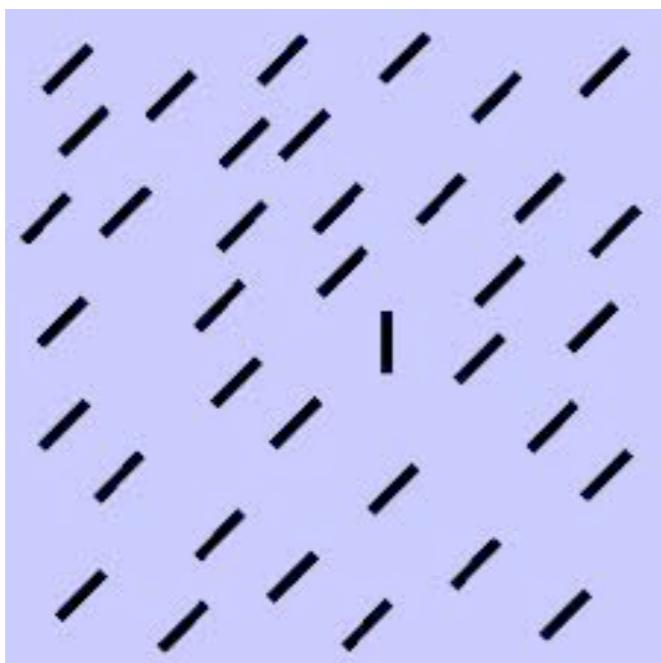
orientation

range-limited

symmetry properties of the
geometry

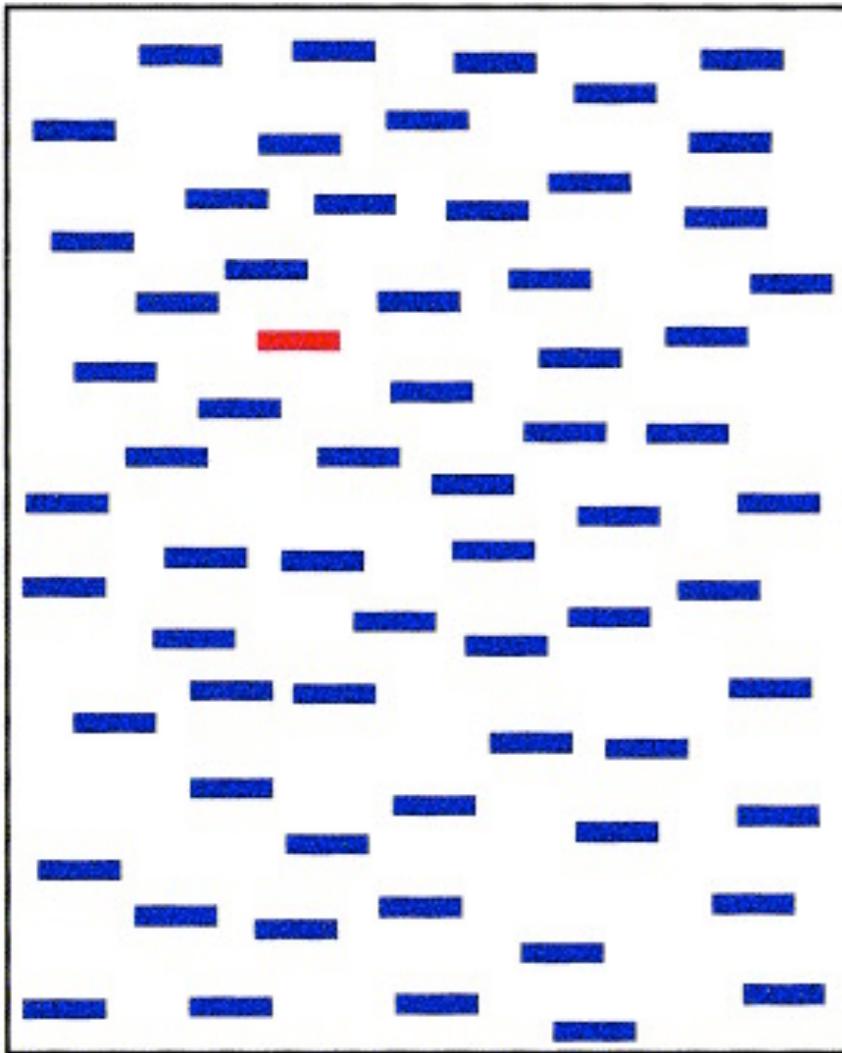


pop-out

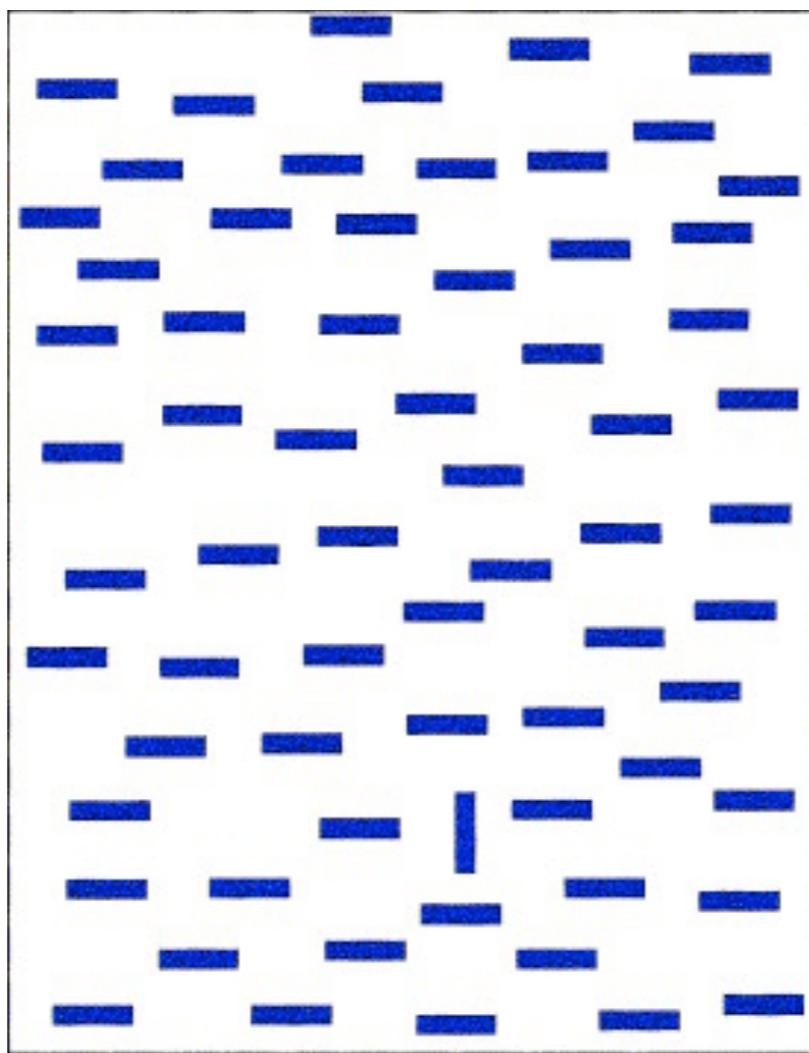


TTTT	FFFF	UUUU
TTTT	F E FF	U <u>U</u> U
TTTT	FFFF	UUUU
TTTT	FFFF	UUUU
TTTT	FFFF	UUUU
T T TT	F A FF	U <u>U</u> U
TTTT	FFFF	UUUU
TTTT	FFFF	UUUU
TTTT	FFFF	UUUU

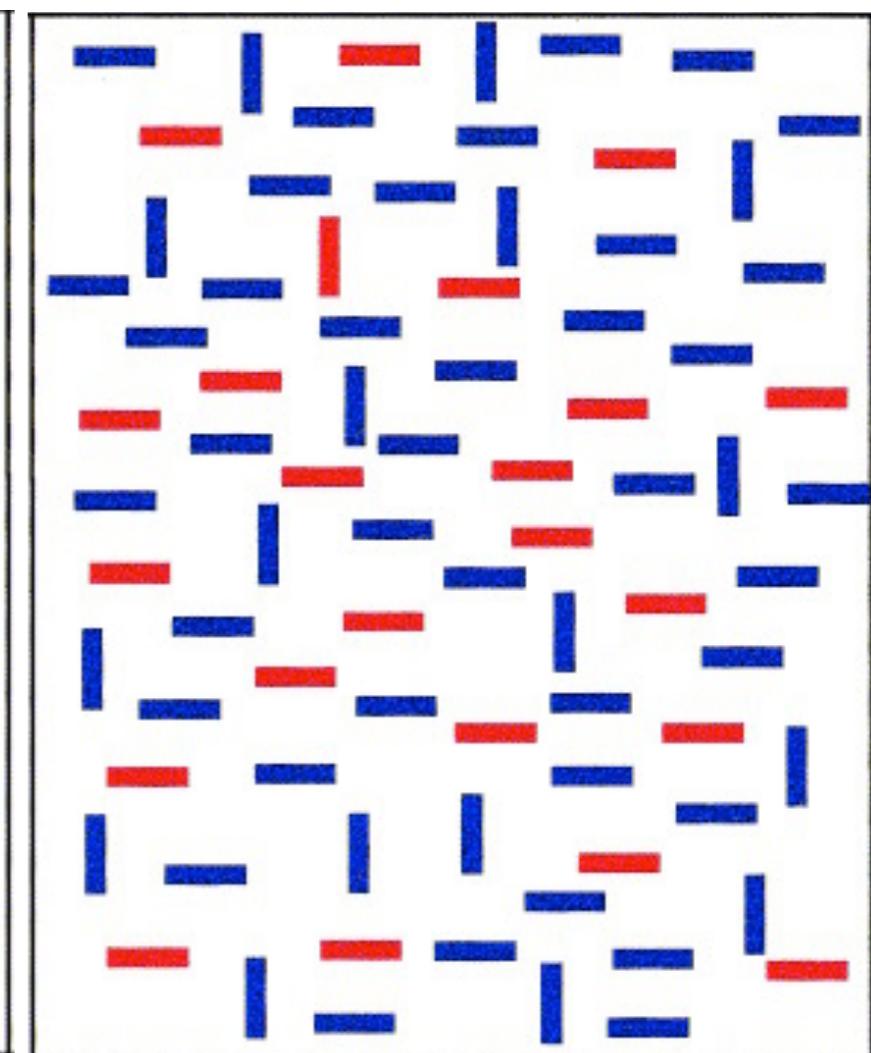
orientation popouts using multiple dimensions



1D colour

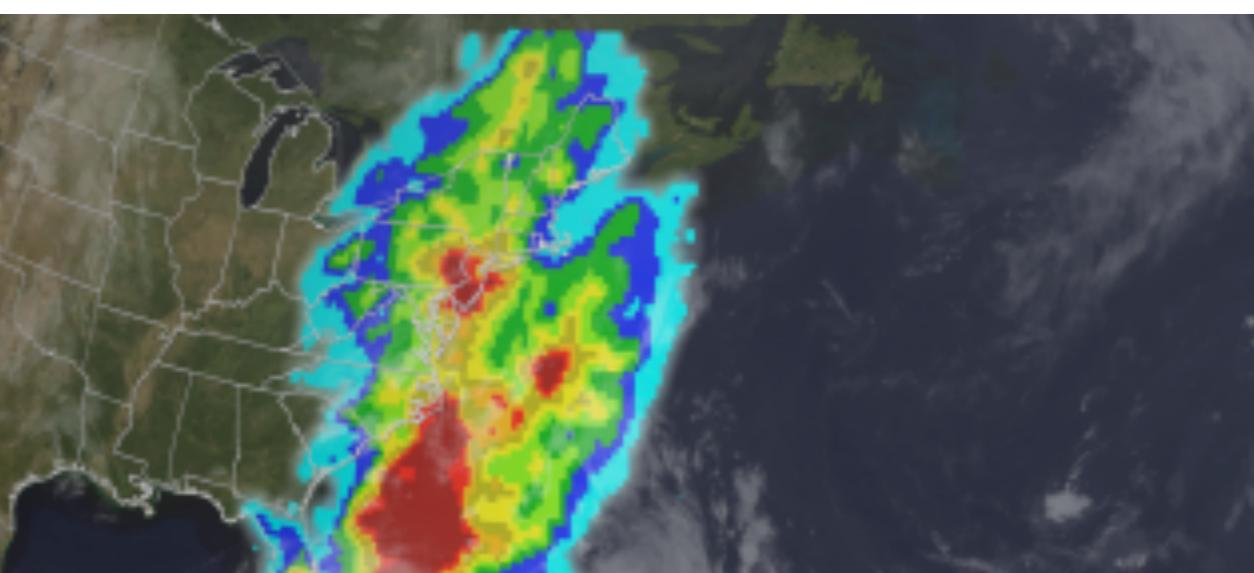
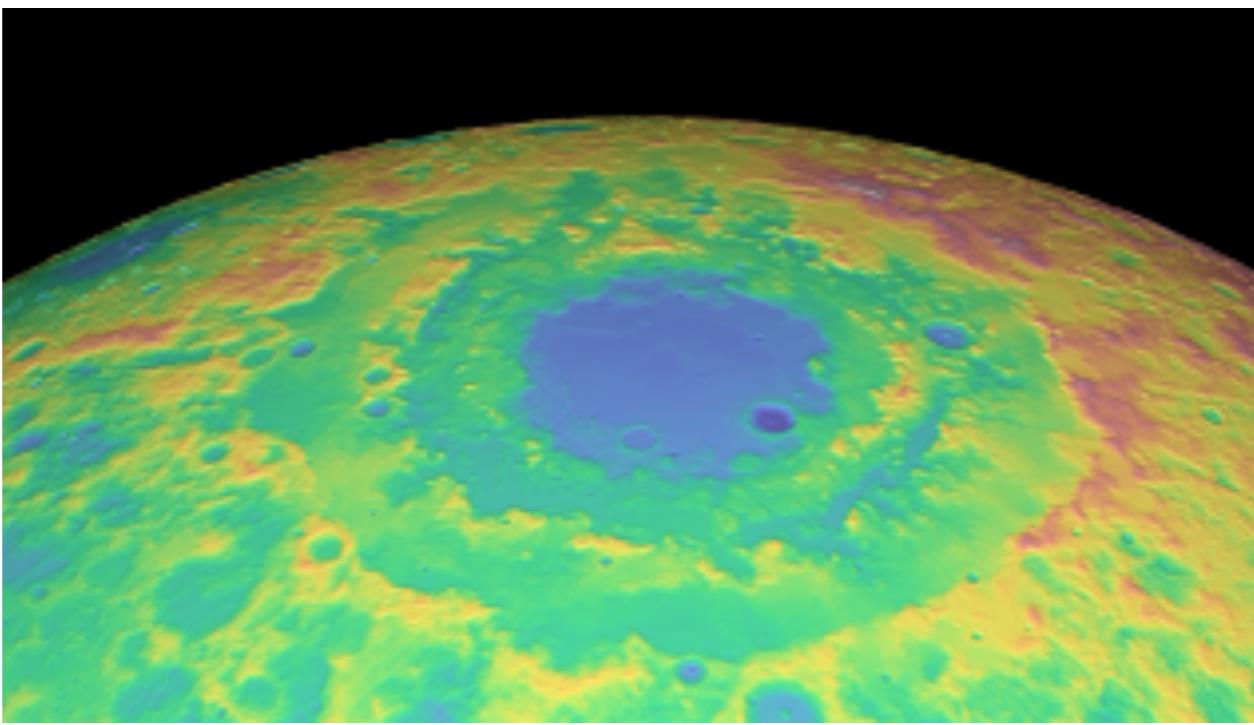
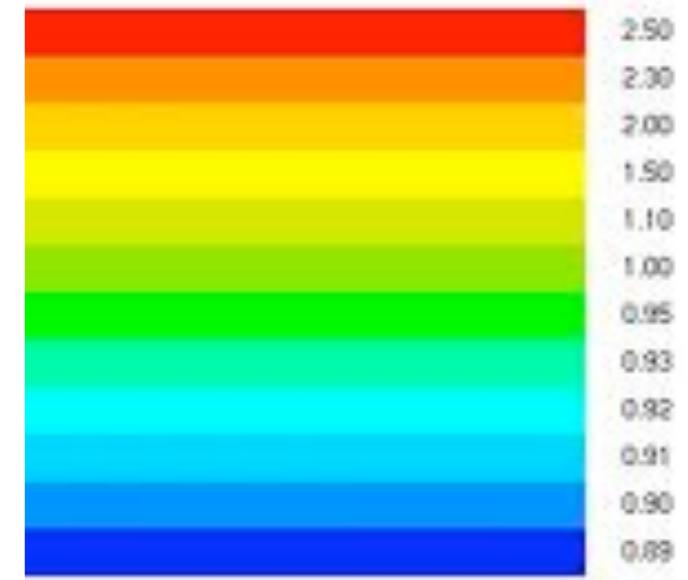
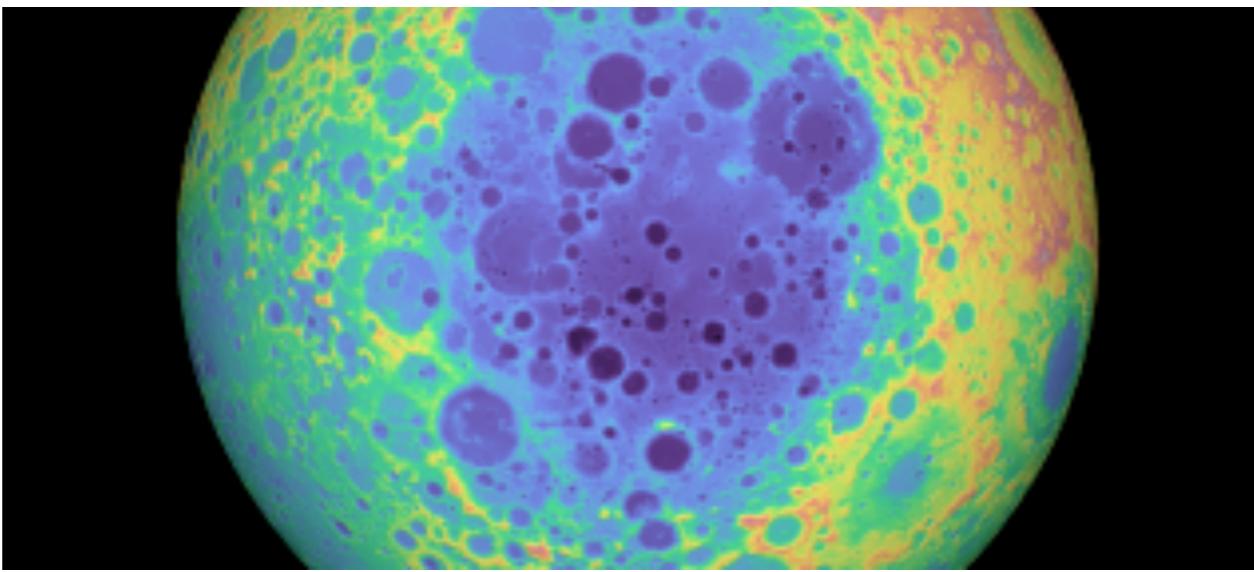
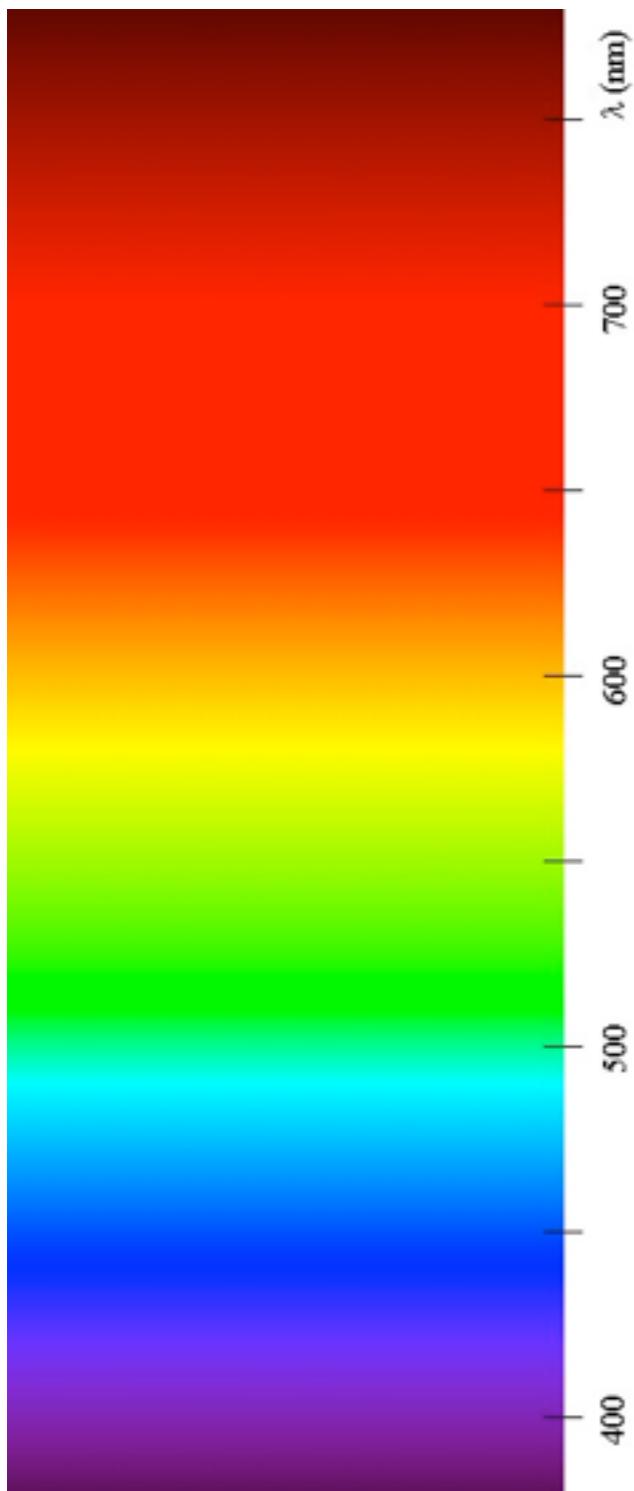


1D orientation



2D color/
orientation

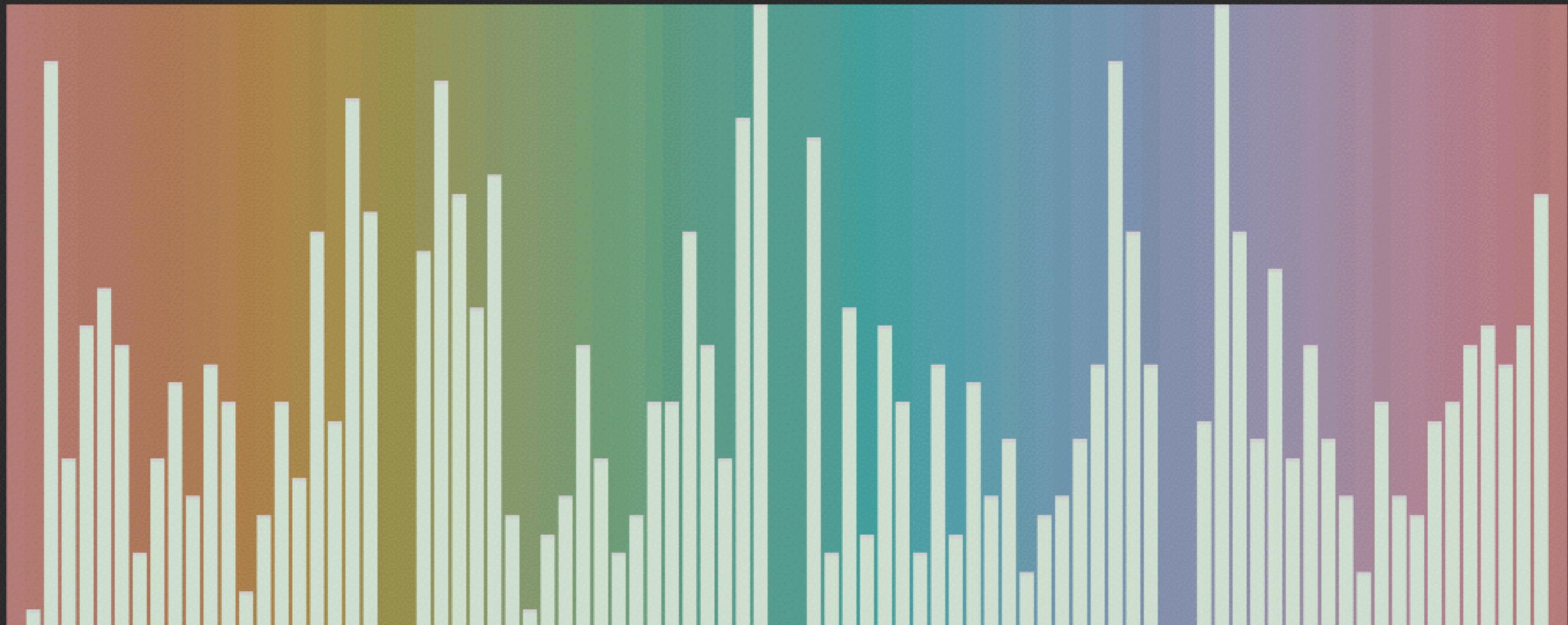
Using colour for continuous values



Using colour for continuous values

Drag and drop the colors in each row to arrange them by hue order.

The first and last color chips are fixed. Click on "Score Test" when done.

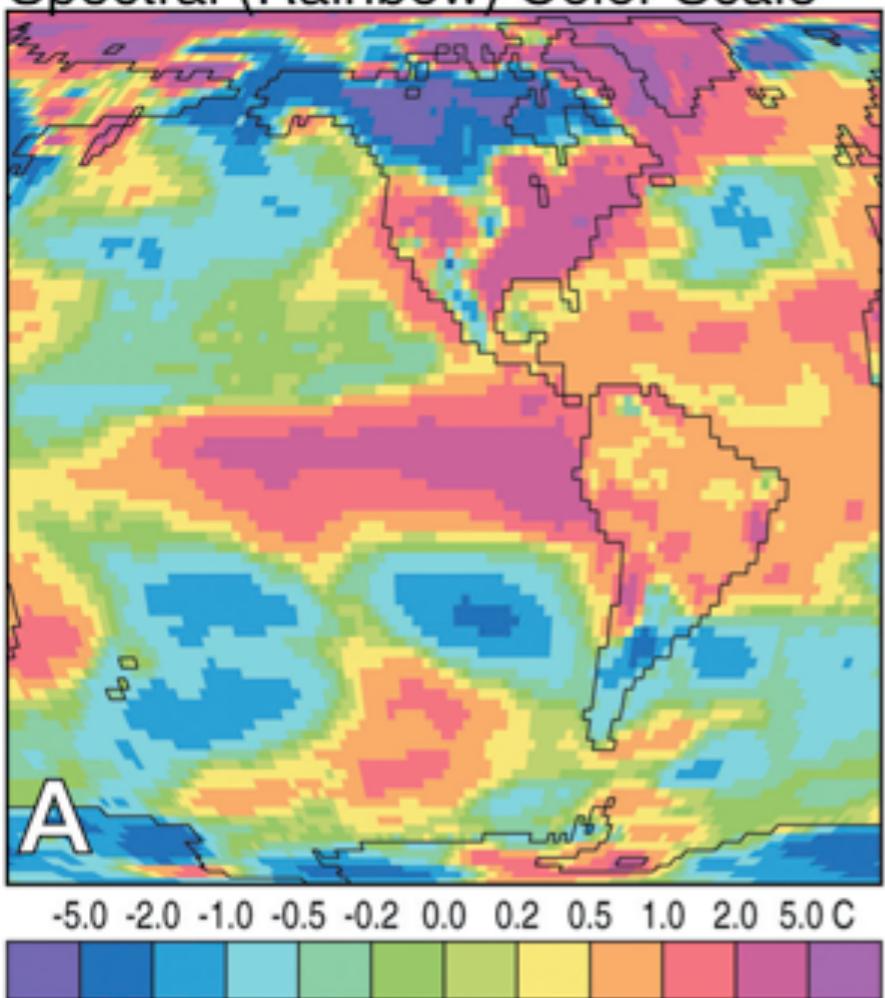


http://www.colormunki.com/game/huetest_kiosk

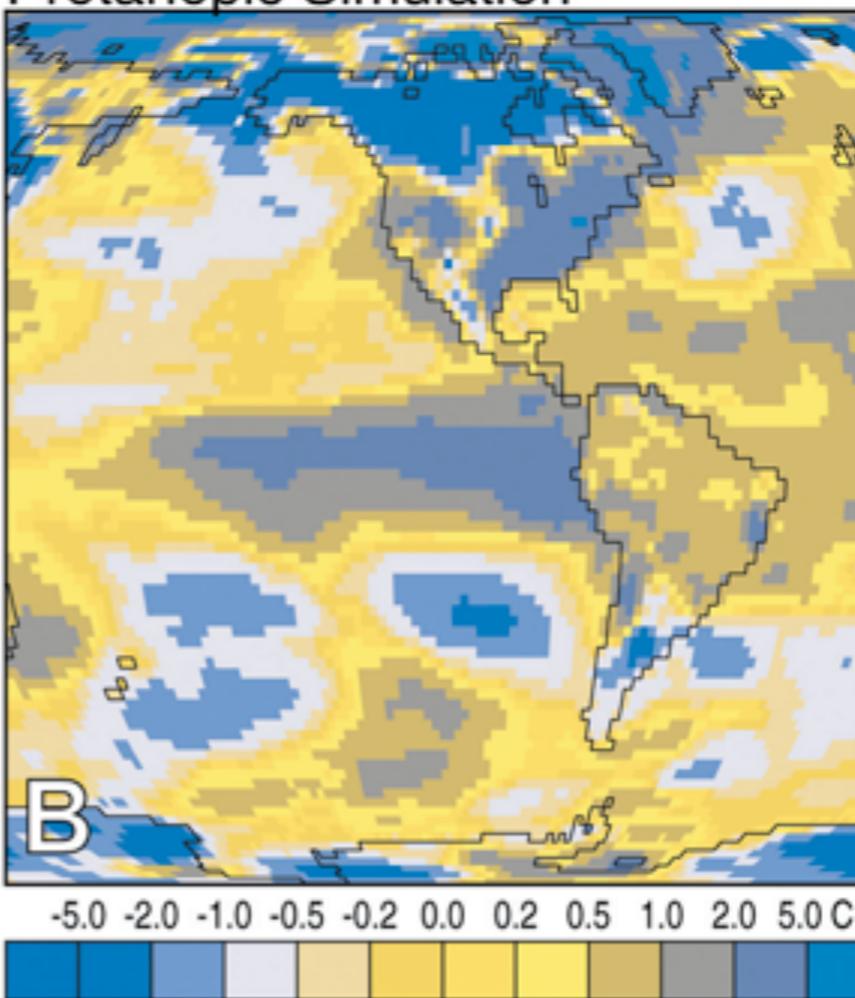
problem 1: No natural ordering

Using colour for continuous values

Spectral (Rainbow) Color Scale

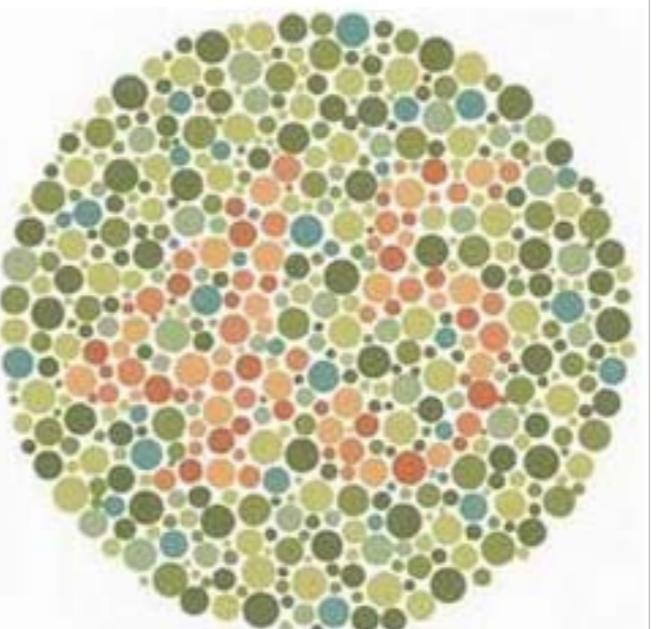
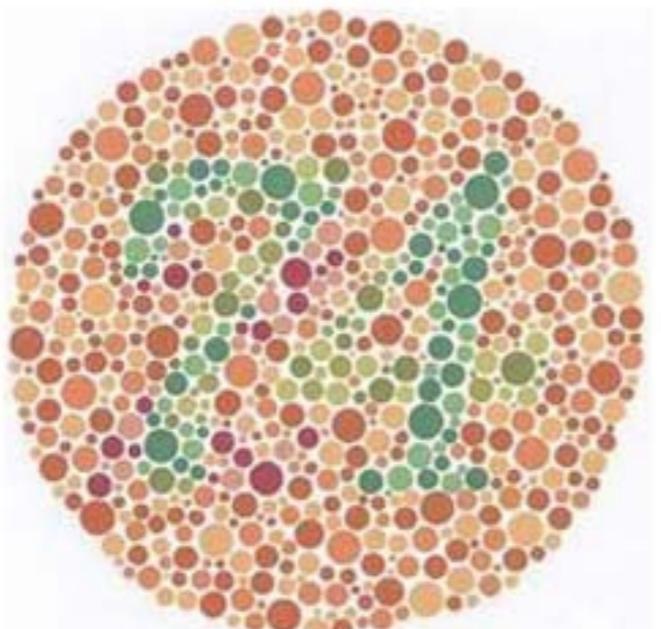
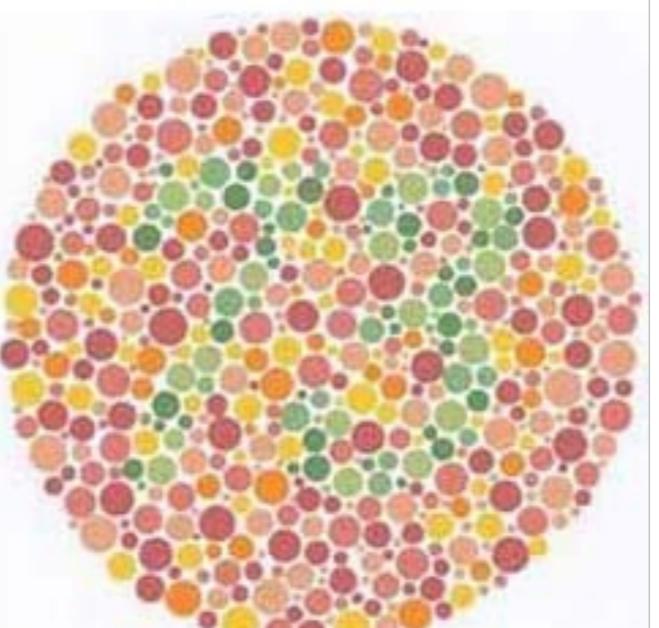
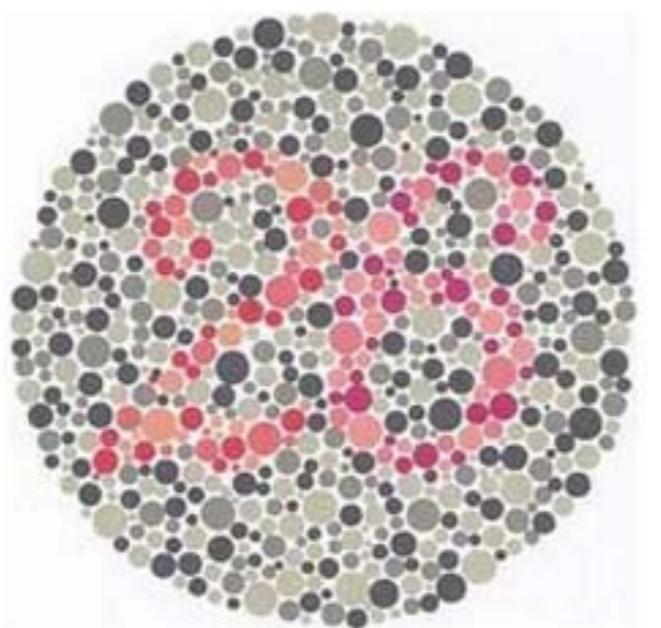
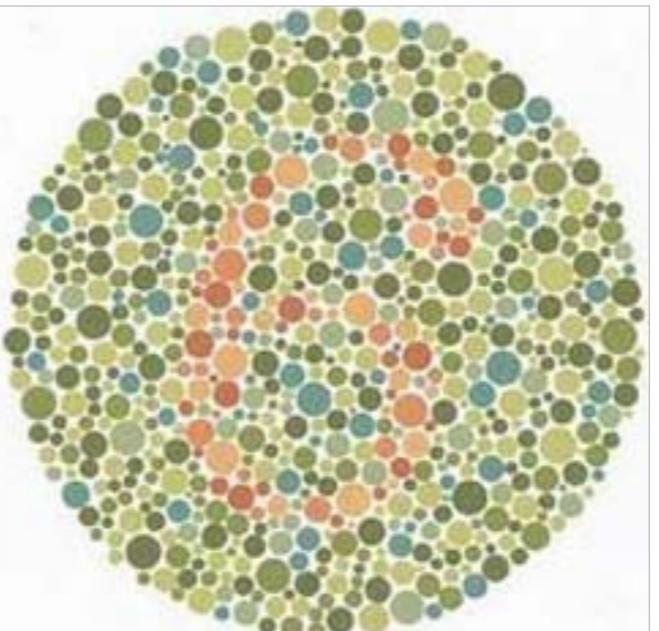
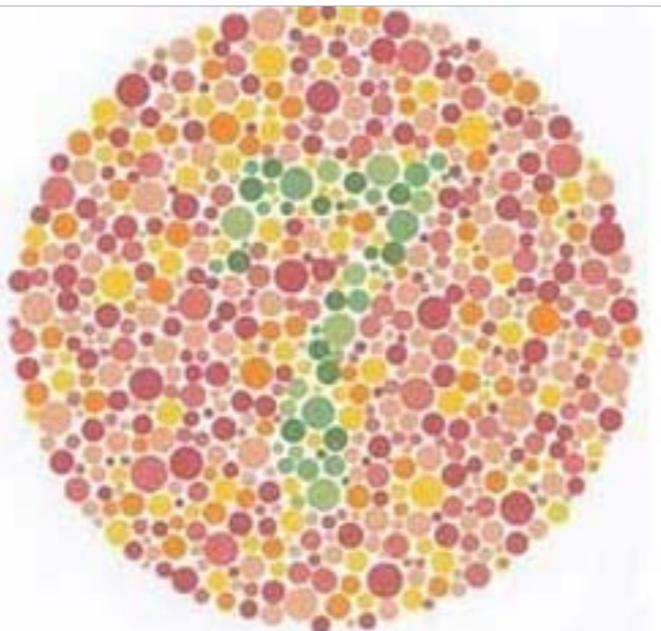


Protanopic Simulation



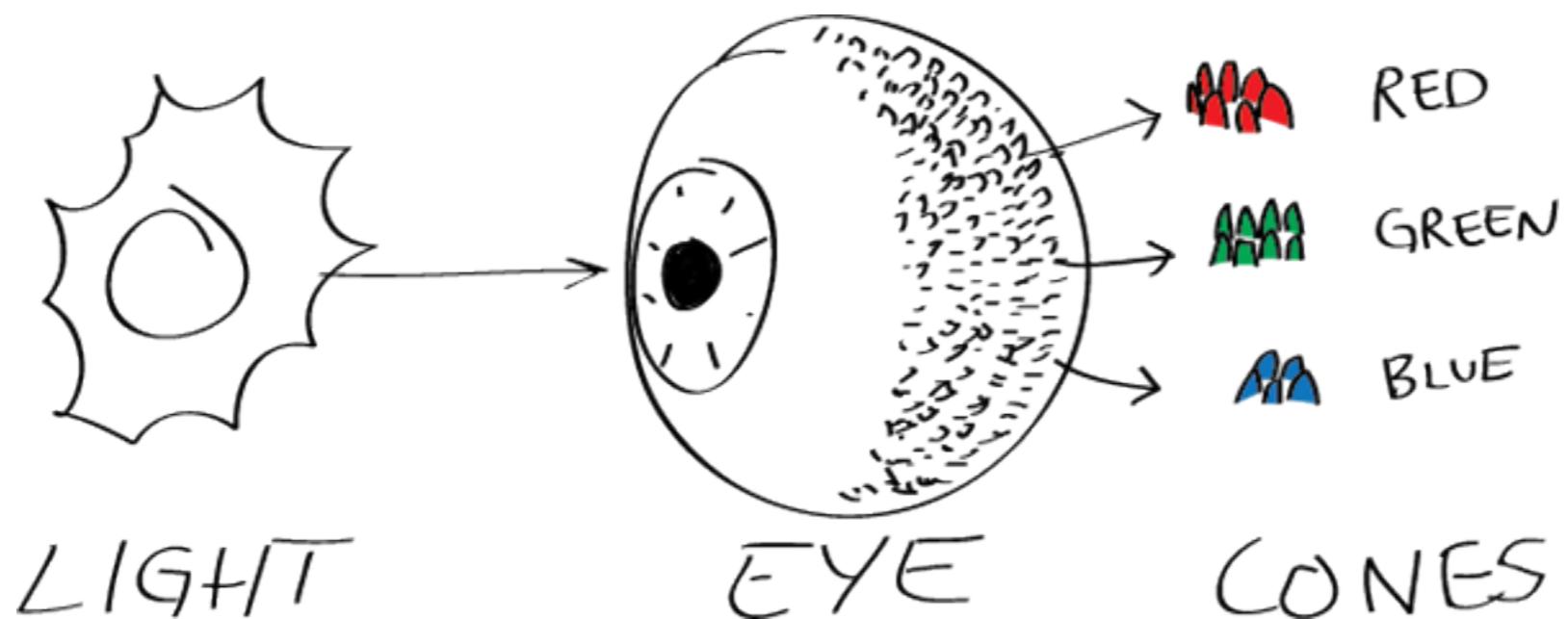
Protanopia affects 8% of males, 0.5% females
of Northern European ancestry

problem 2: **colour** sensitivity

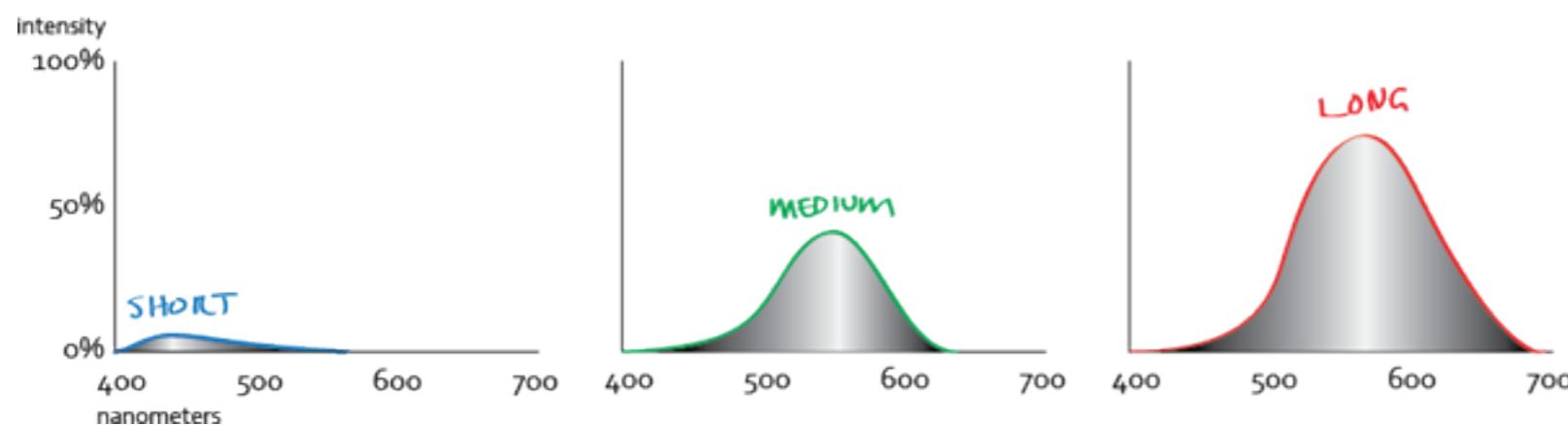


Using colour for continuous values

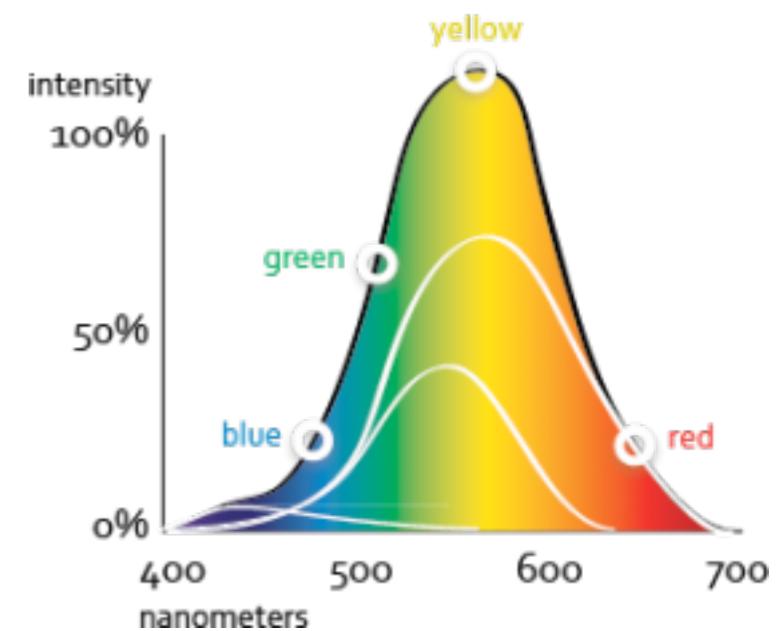
problem 3: yellow is special



RELATIVE SENSITIVITY TO LIGHT WAVELENGTHS

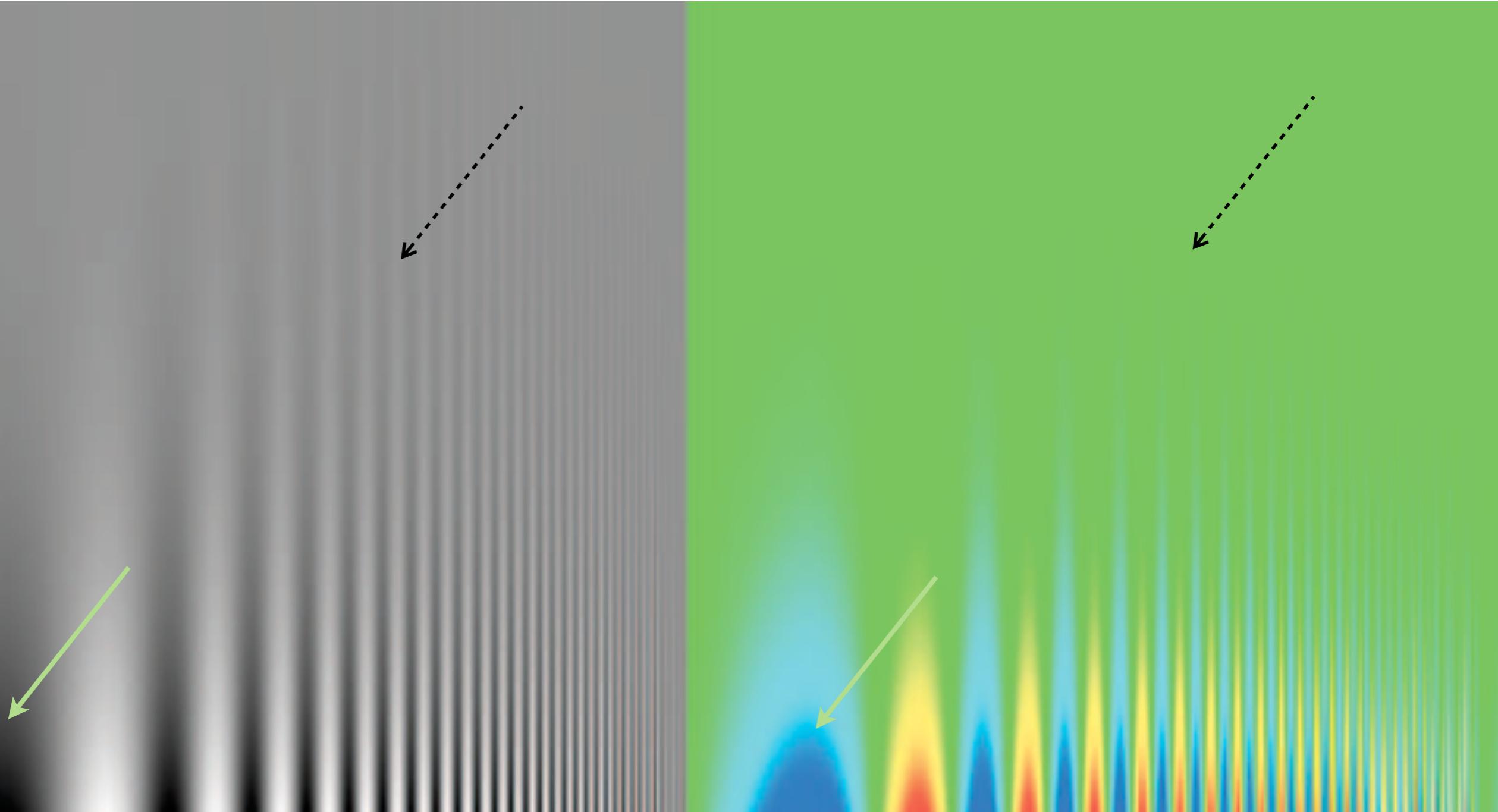


PUTTING IT ALL TOGETHER



Using colour for continuous values

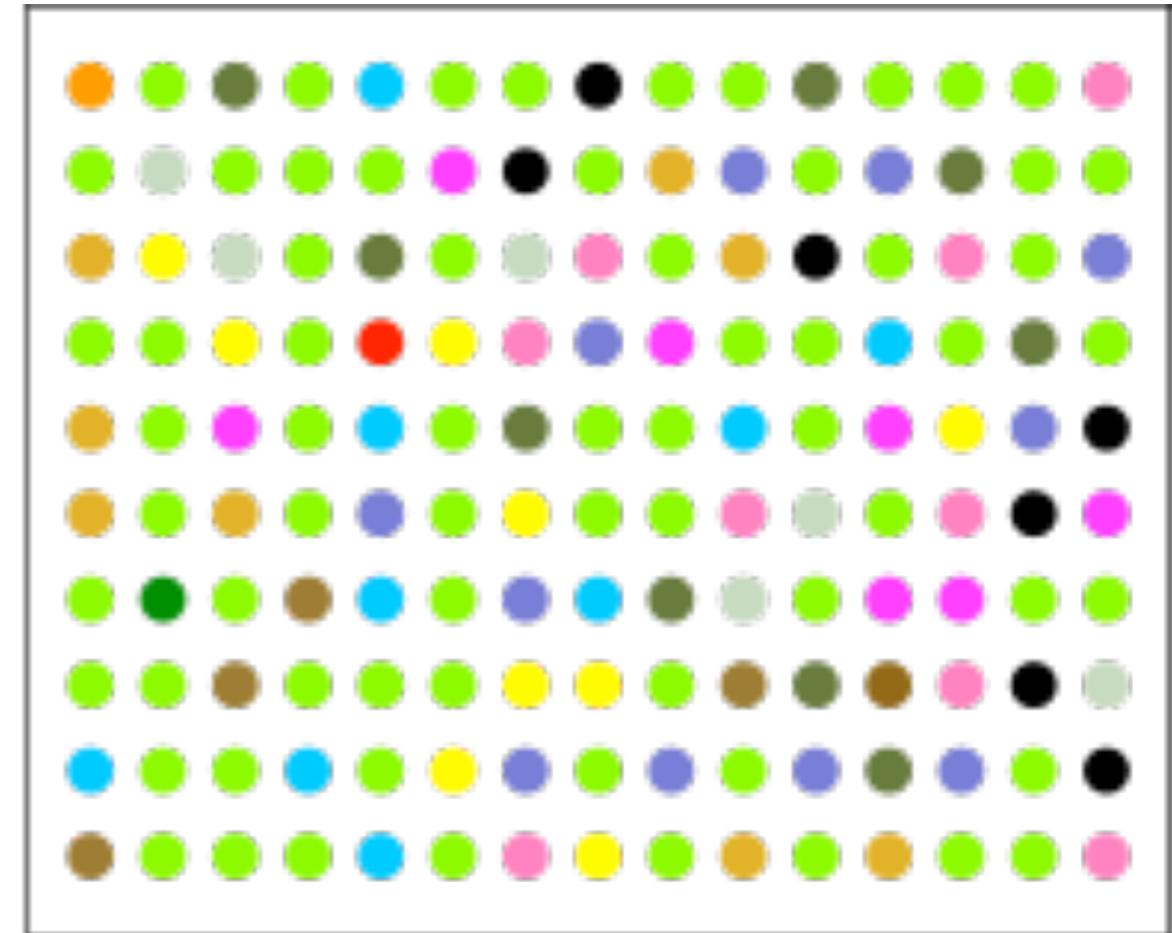
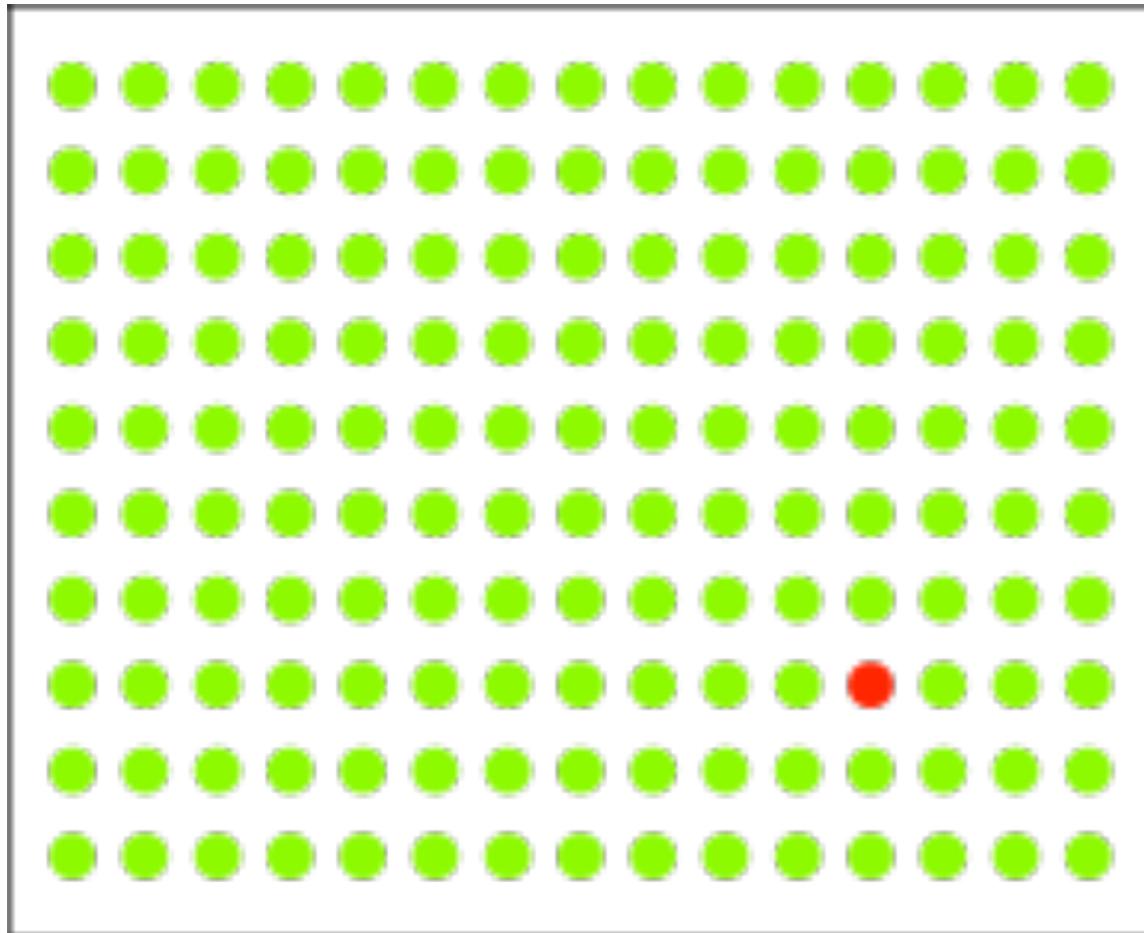
problem 4: Details: overemphasised or obscured

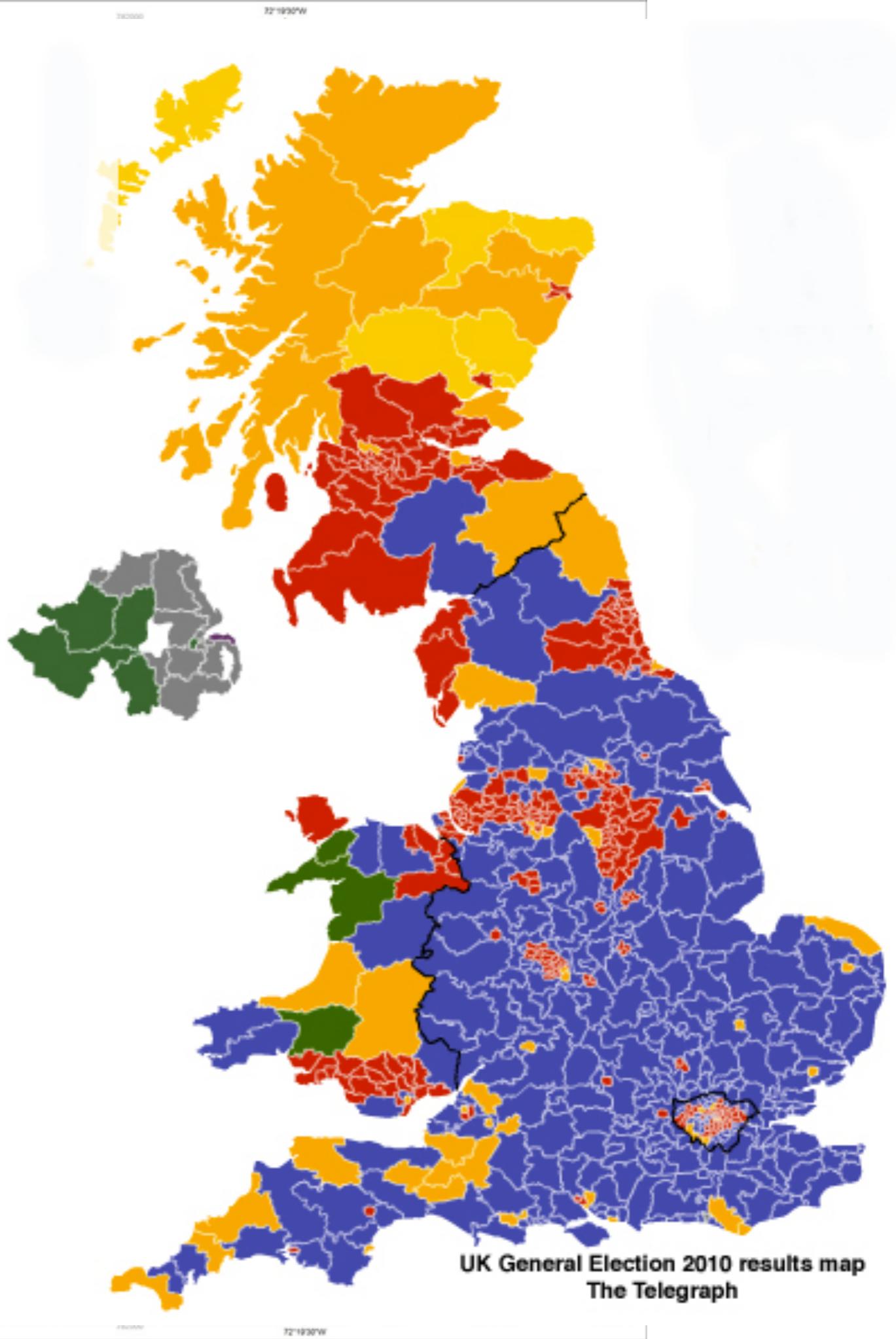
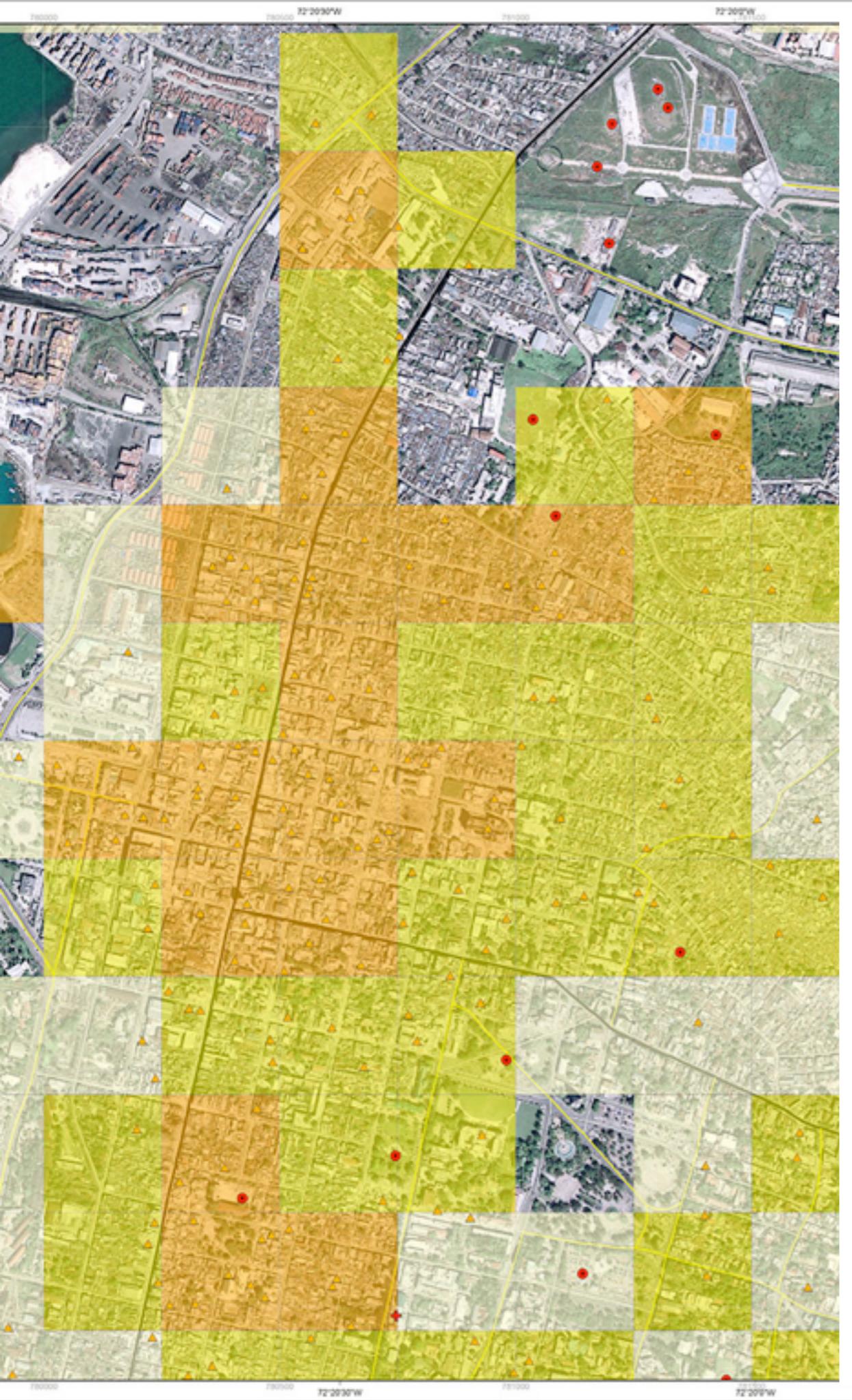


hue ‘borders’ overemphasise small changes, hue ‘middles’ blend potentially important details

Using colour for continuous values

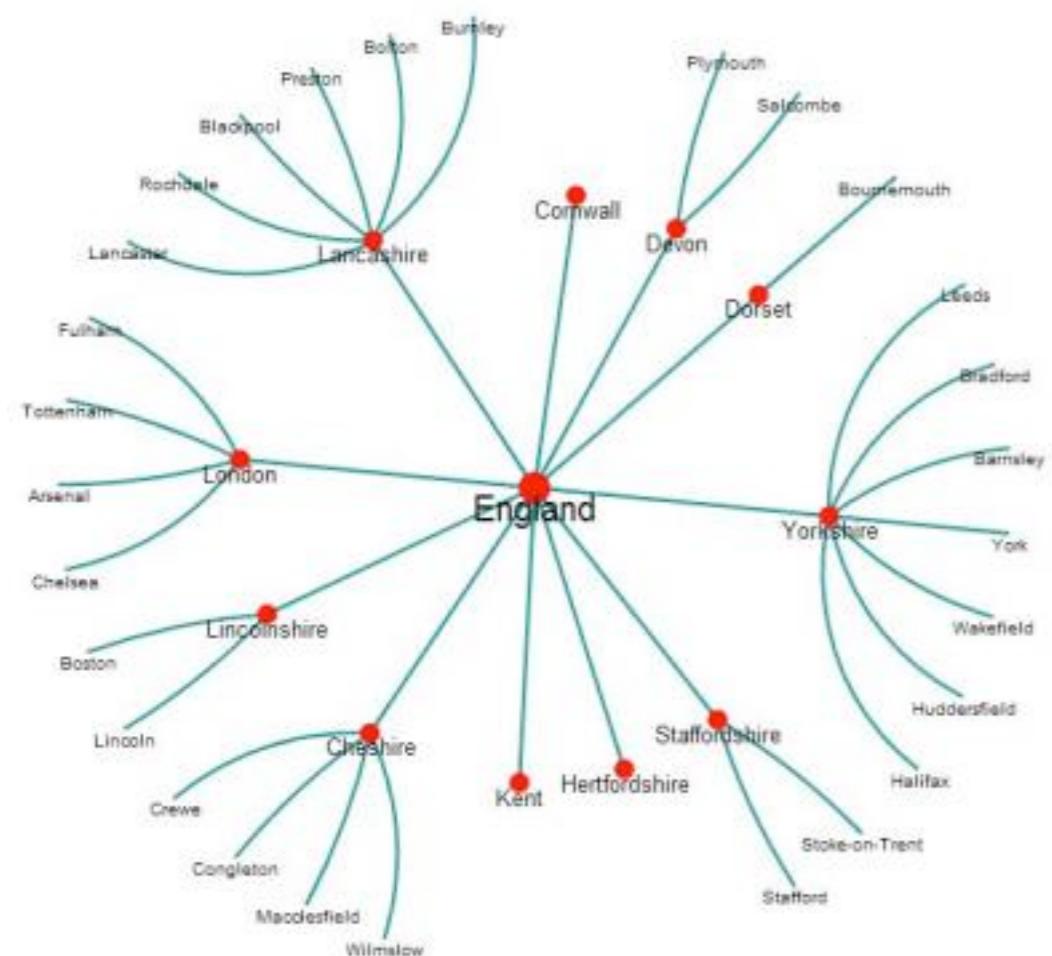
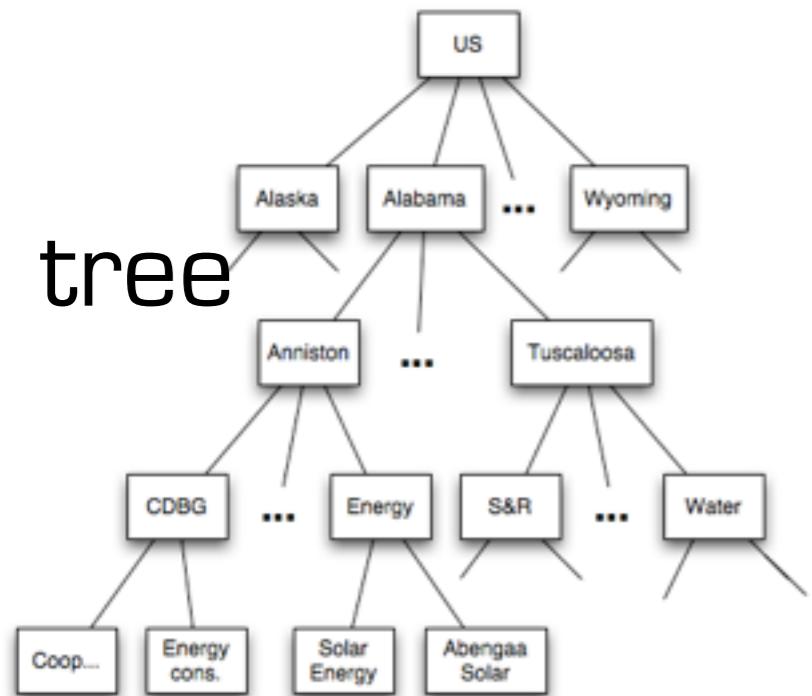
problem 5: pop out can drown out



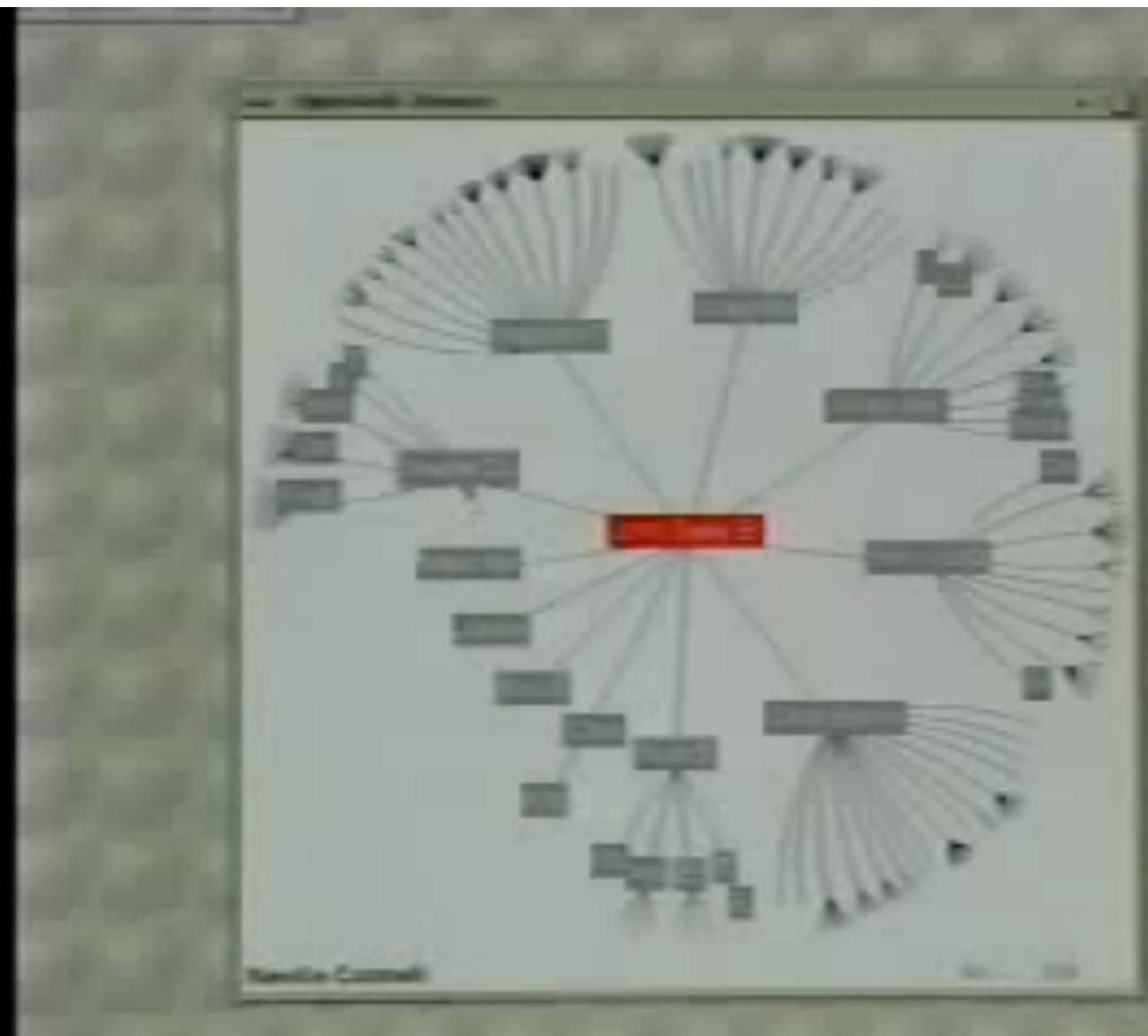


multivariate relational data: hierarchical

tree

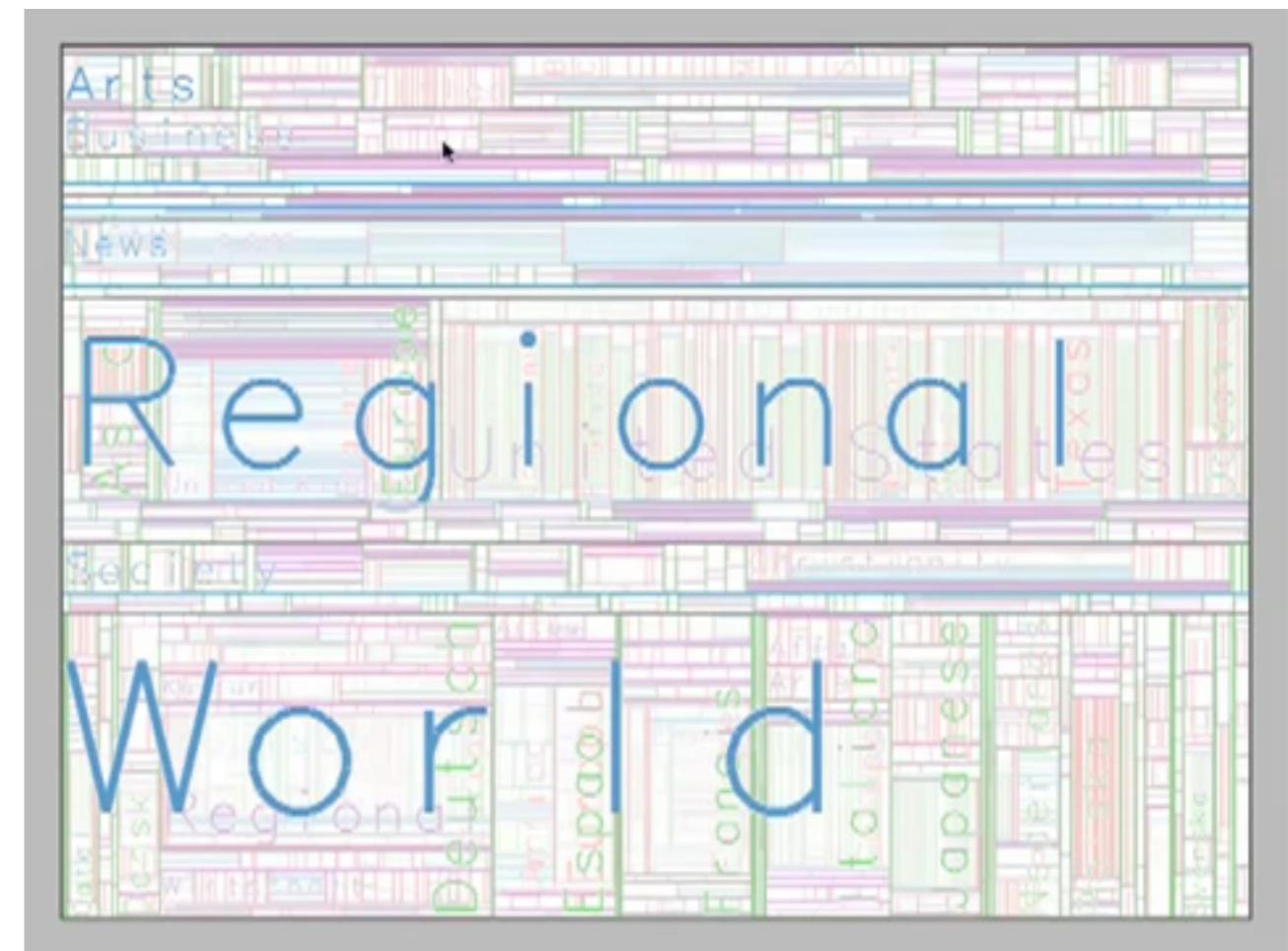
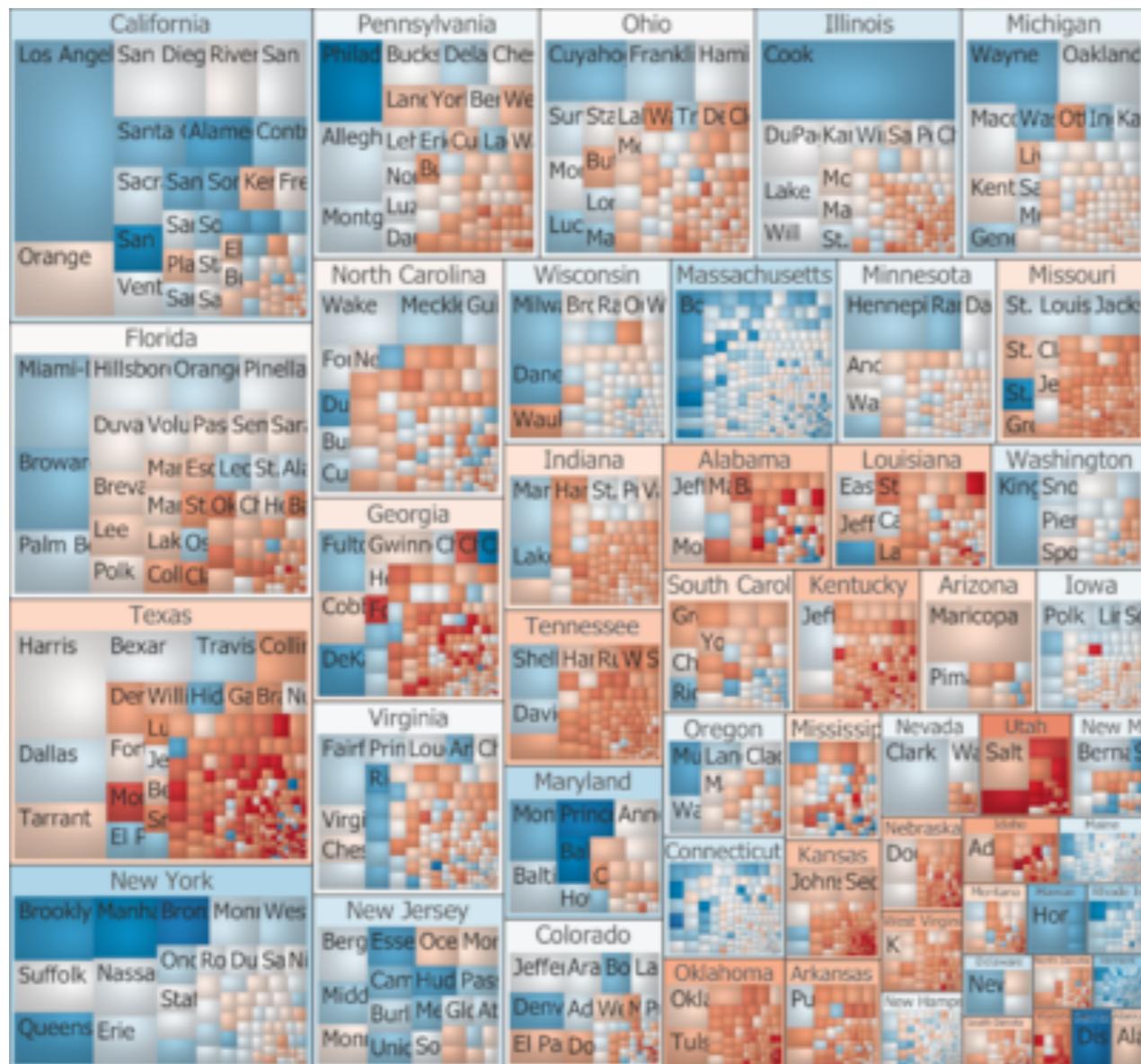


hyperbolic tree

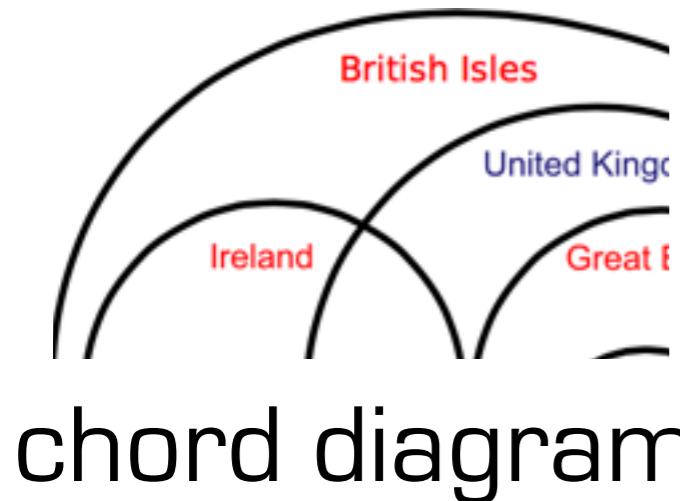
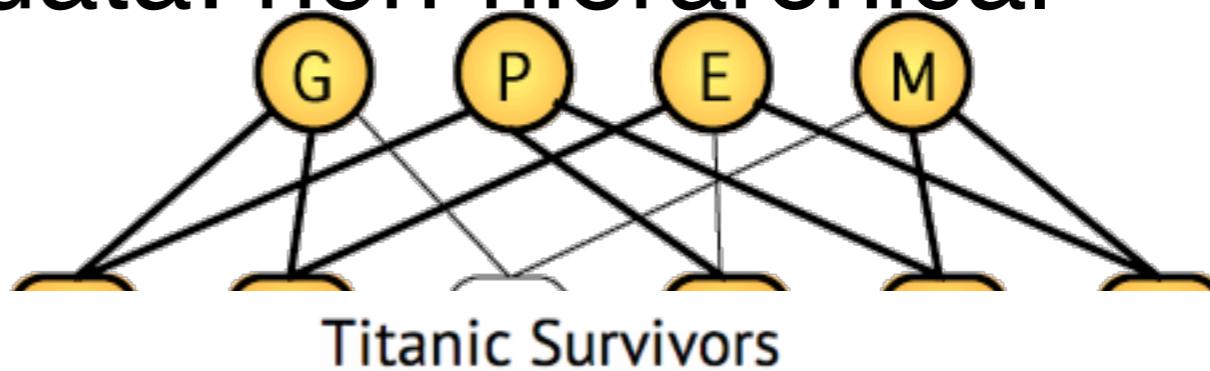


multivariate relational data: hierarchical

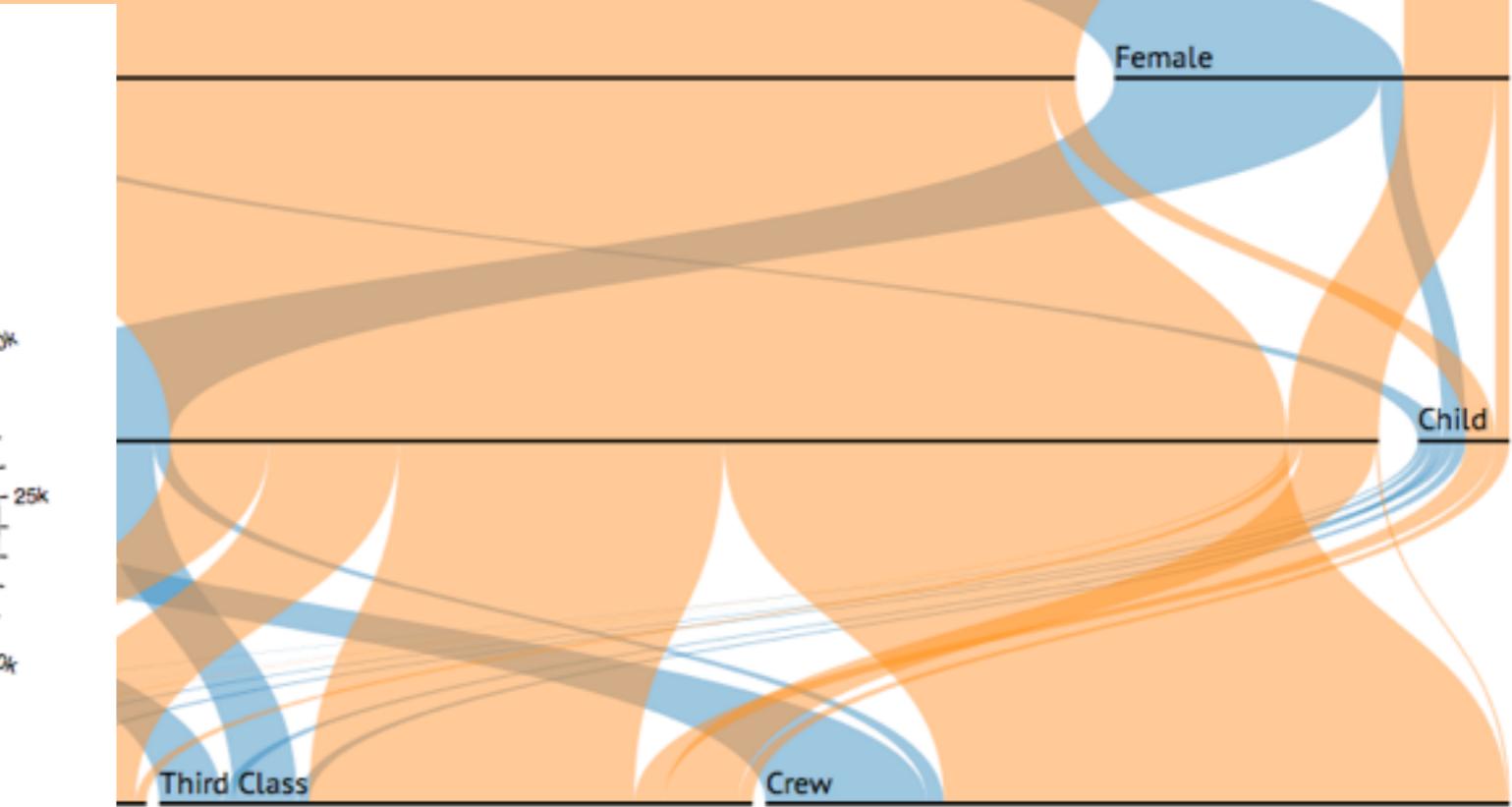
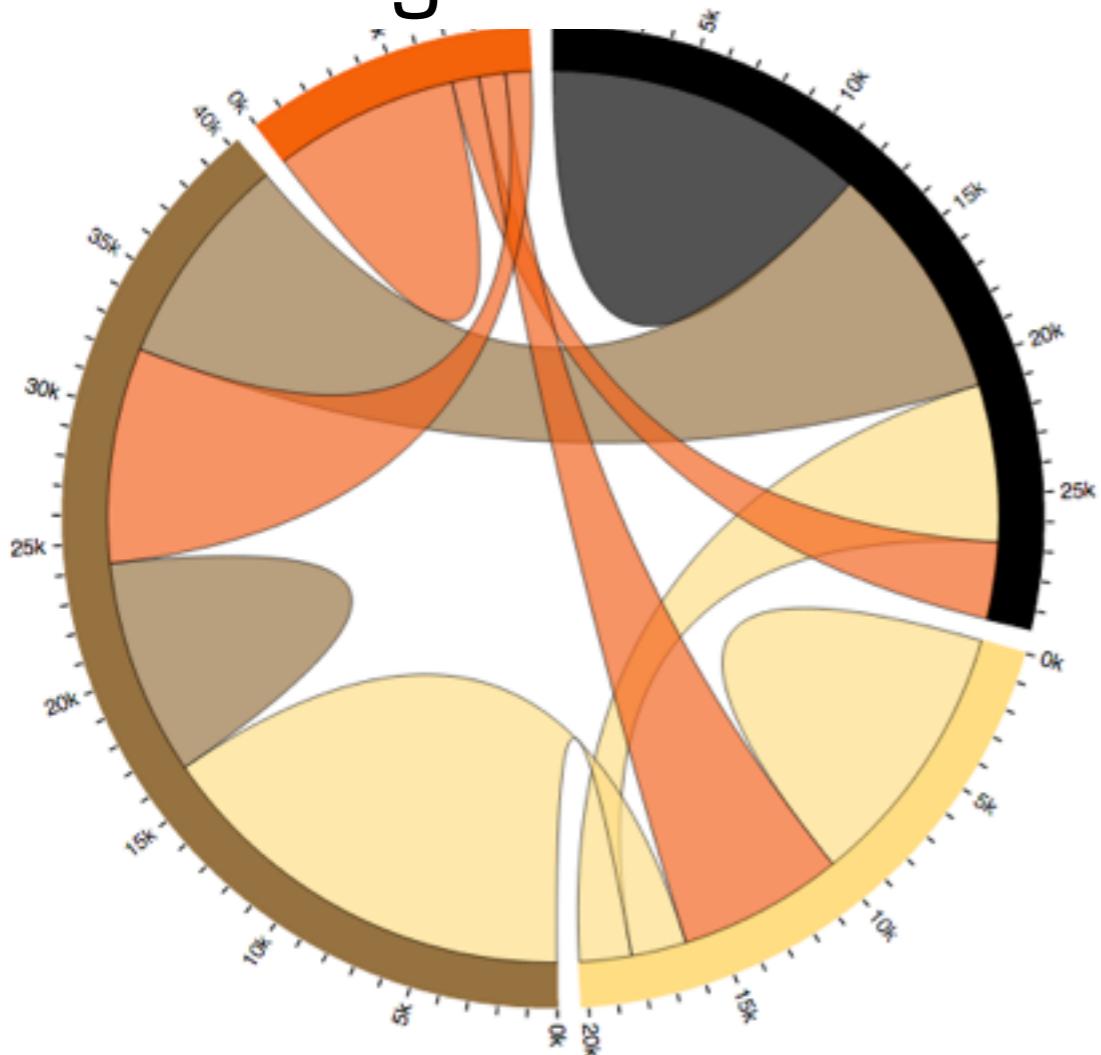
treemap



multivariate relational data: non-hierarchical



chord diagram



Data: [Robert J. MacG. Dawson](#).

;S

time series (animation)



aaron koblin - flight patterns

time series (static) - small multiples

Income under \$20,000



\$20-40,000



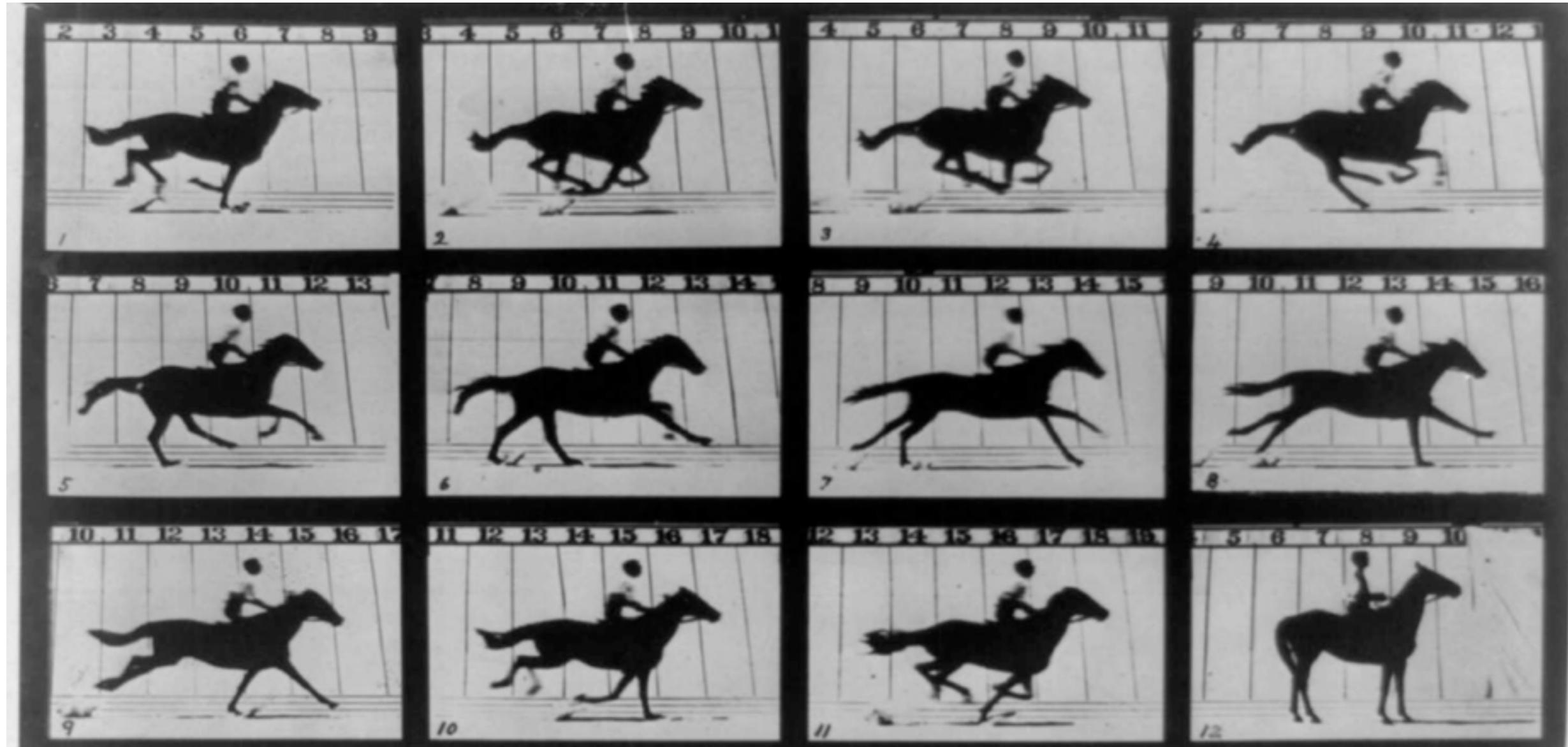
\$40-75,000



\$75-150,000



Over \$150,000



Copyright, 1878, by MUYBRIDGE.

MORSE'S GALLERY, 417 Montgomery St., San Francisco

THE HORSE IN MOTION.

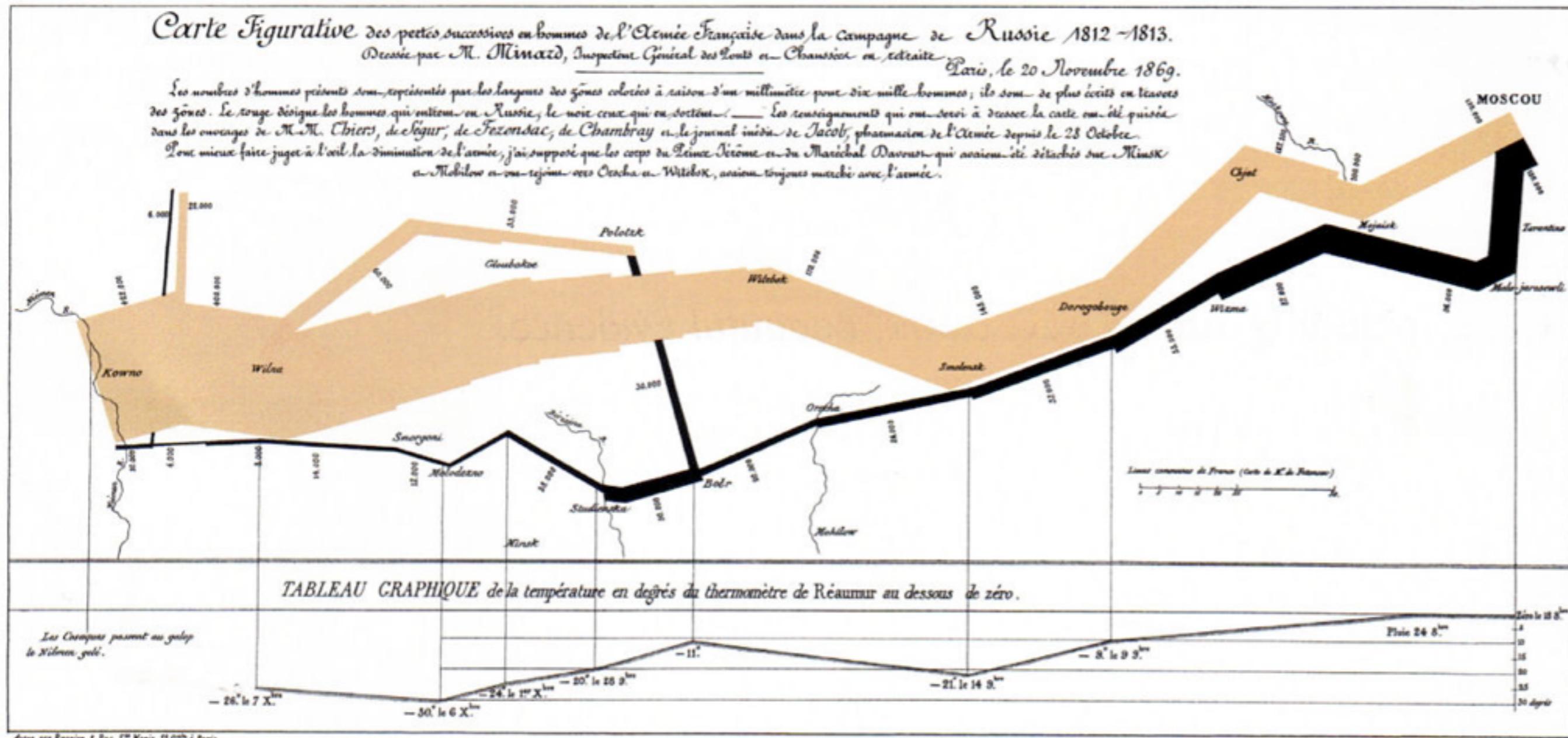
Illustrated by

MUYBRIDGE

Patent for Reproduction Applied for.

AUTOMATIC ELECTRO-PHOTOGRAPHIC

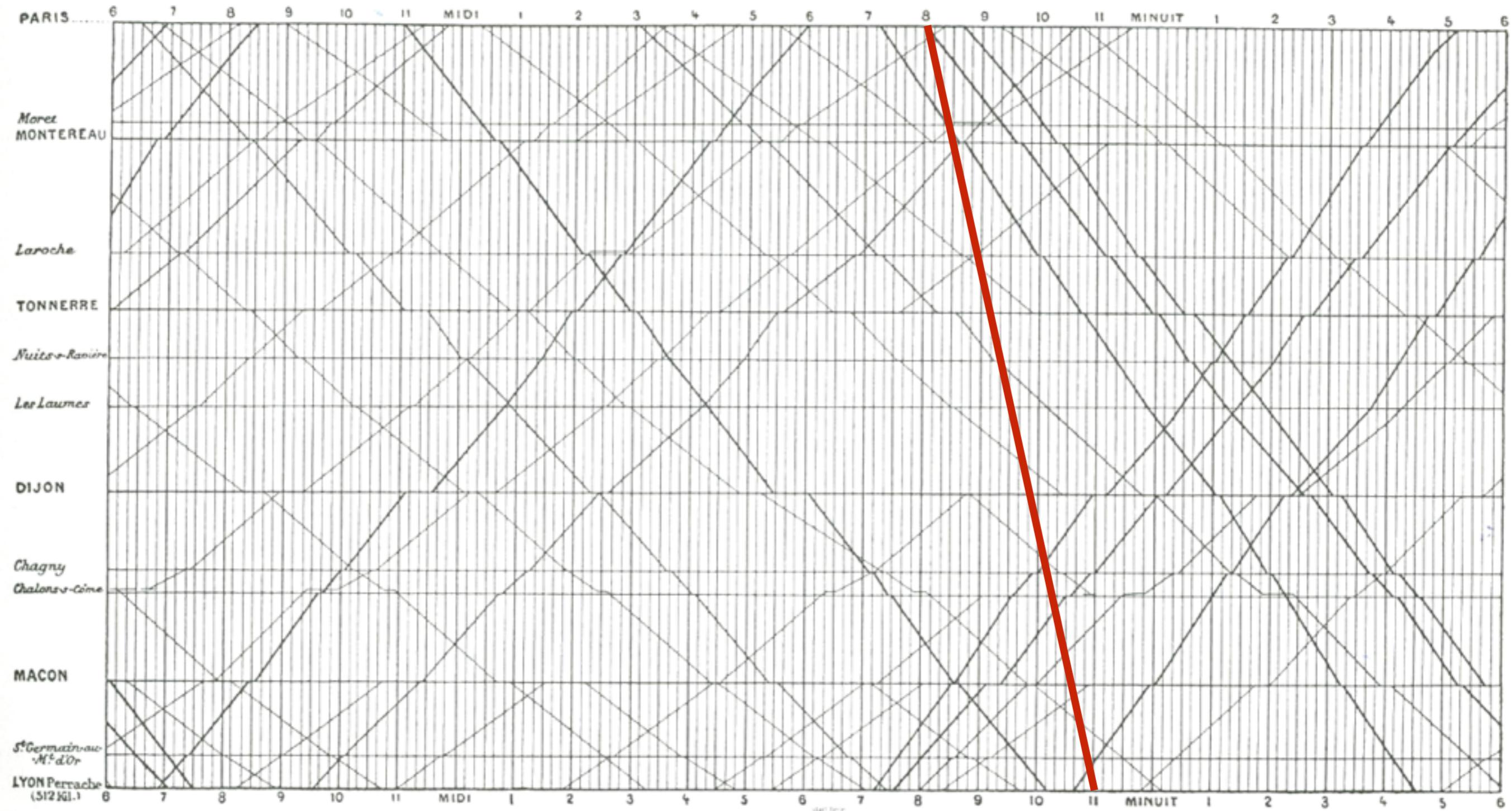
charles joseph minard napoleon's march to moscow (1869)



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

TGV

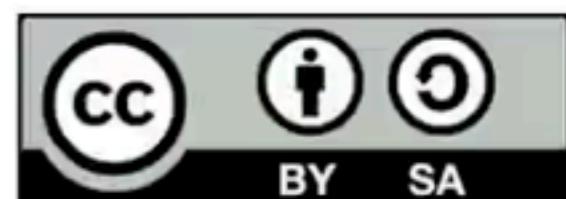


E.J. Marey
La méthode graphique
(1885)

200 years that changed the world

with Hans Rosling

Free to redistribute



www.gapminder.org

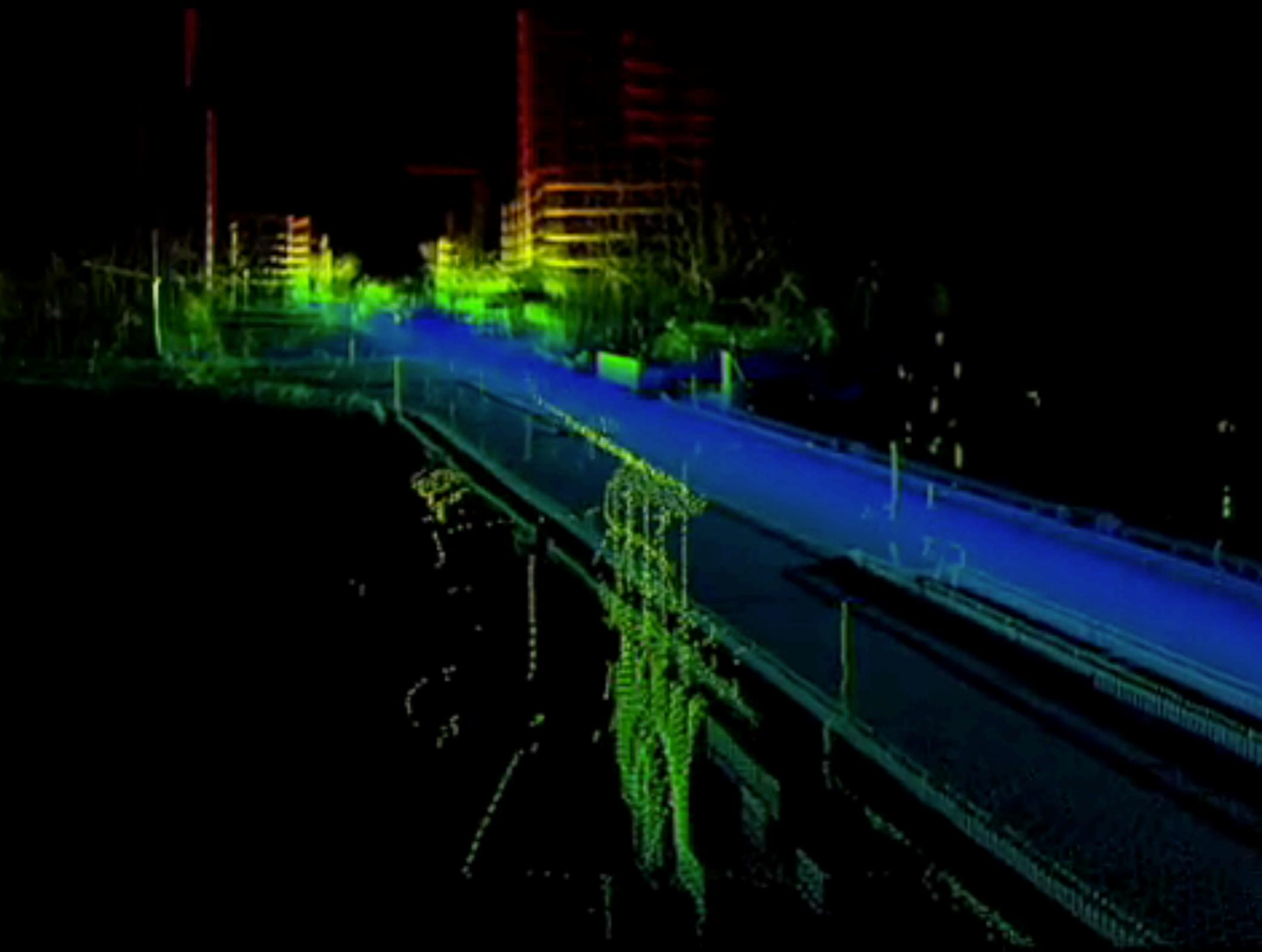
In conclusion

Designing effective infographics

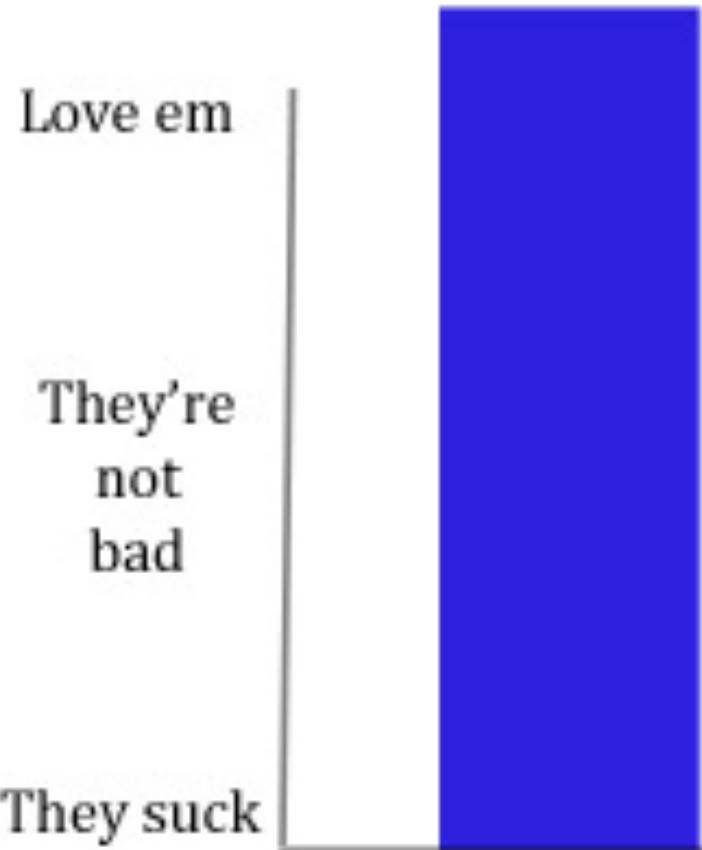
is about effectively conveying or facilitating an understanding of relationships in data

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

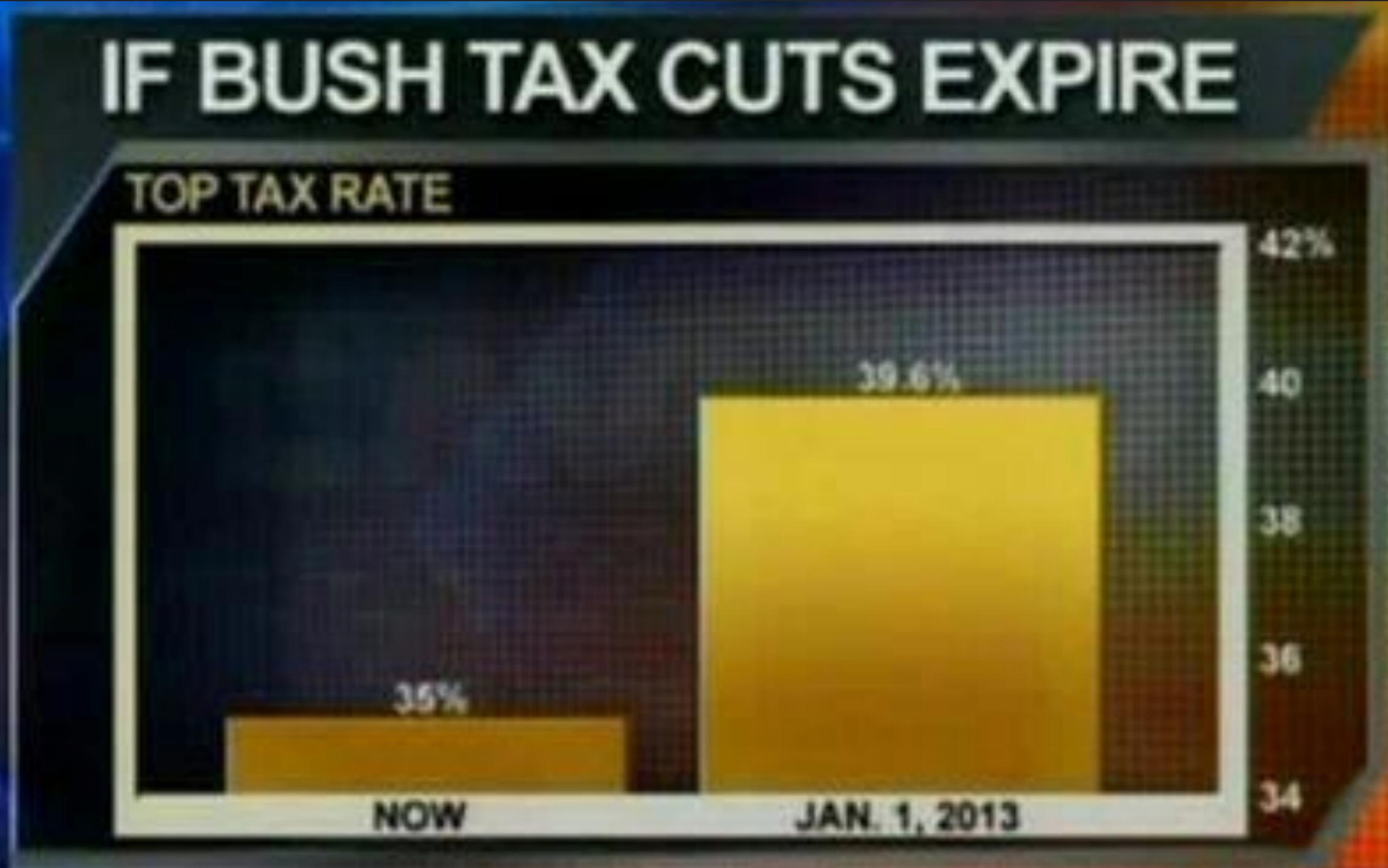


How You Feel About Bar Charts



communicating
through infographics:
visual + statistical sleight of hand to
mislead the audience

1. Barchart baseline fail



8:01 p ET

FOX
BUSINESS

TOP STORIES

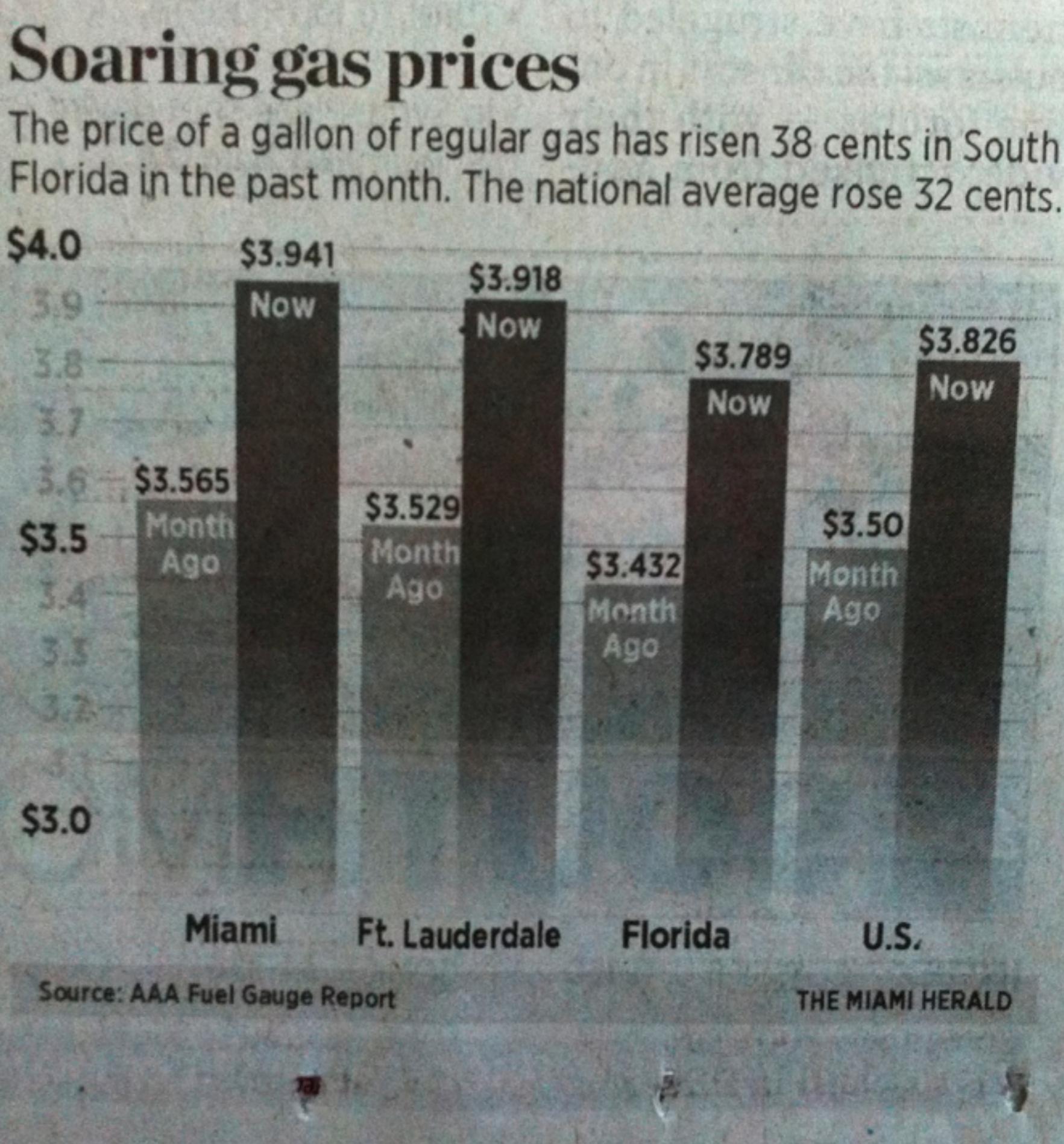
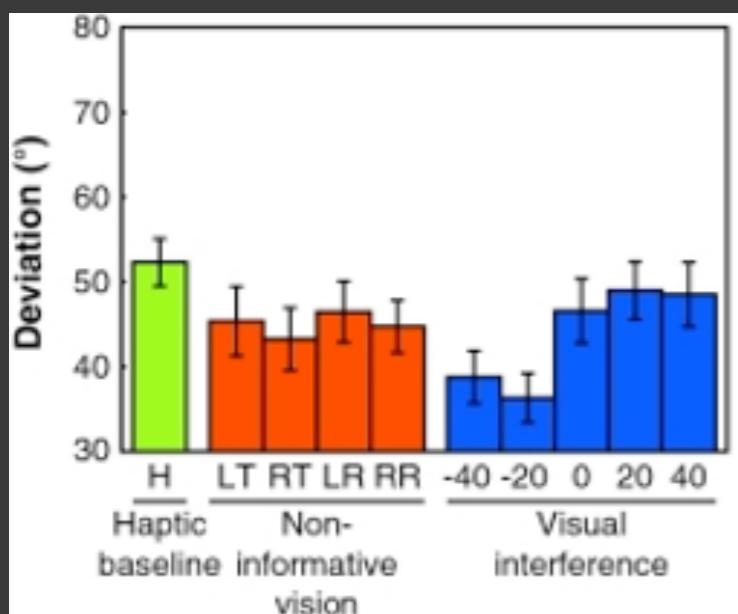
TECHNOLOGY

CONSUMER

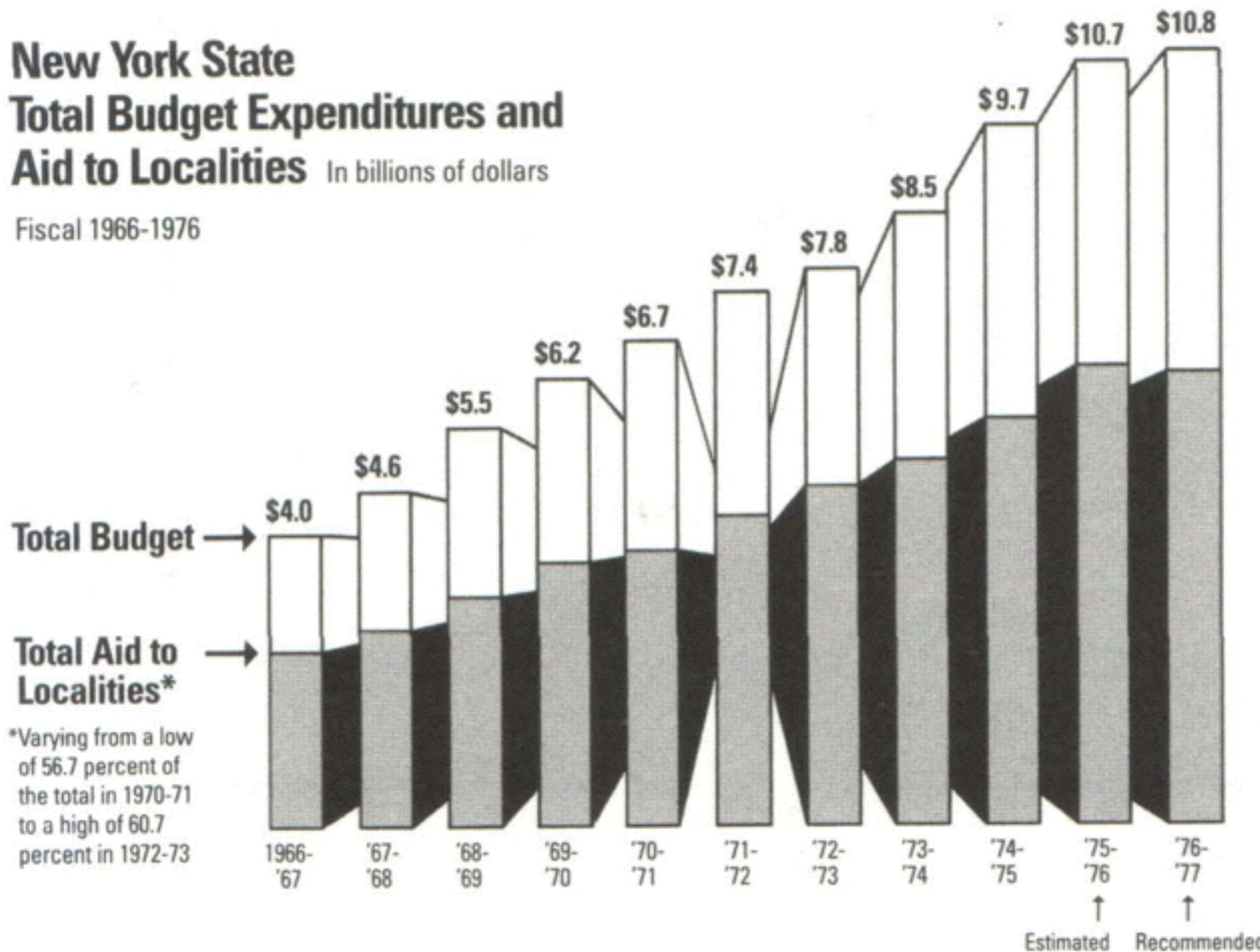
WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

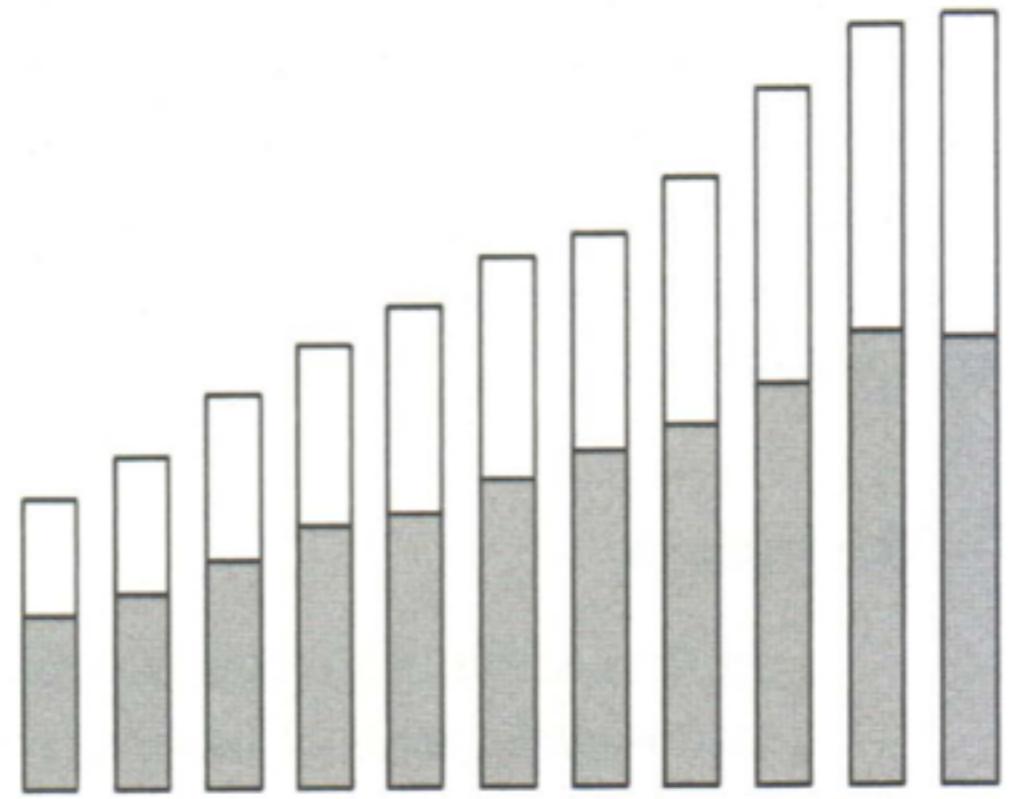
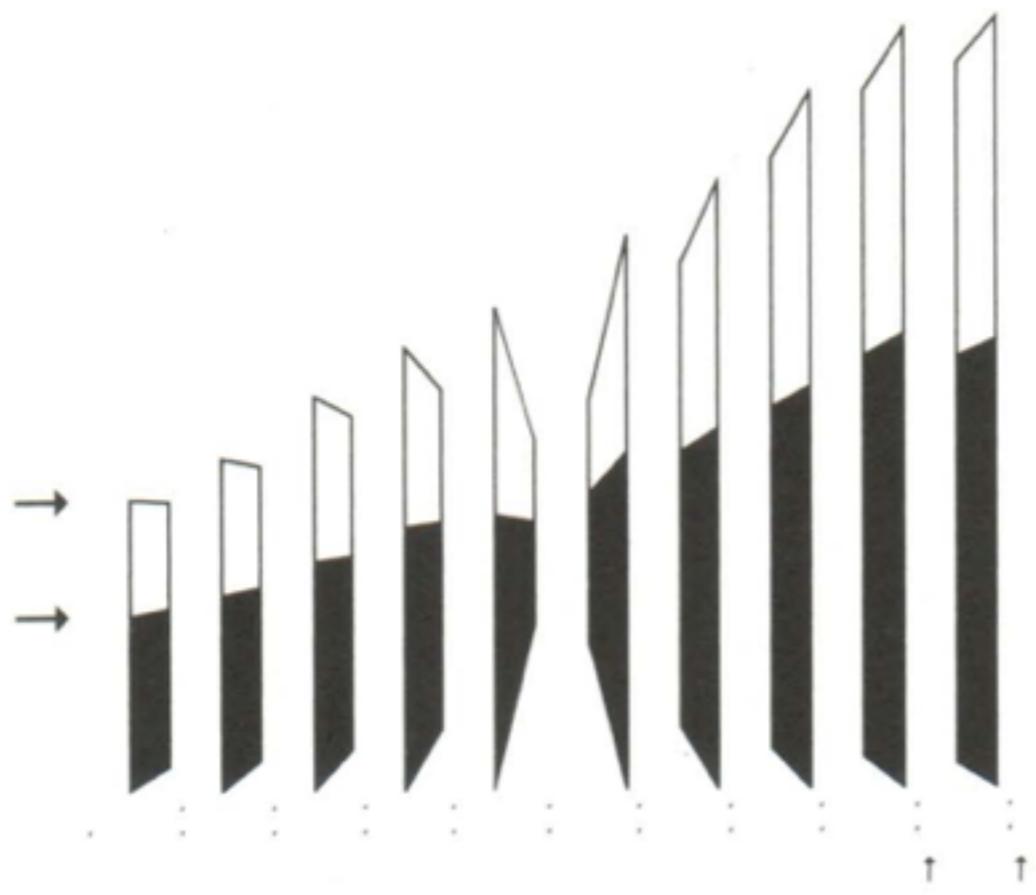
DOW 13008.68 ▲ 64.33 S&P 1379.32 ▲ 5.98 NASDAQ 2939.52 ▲ 6.32

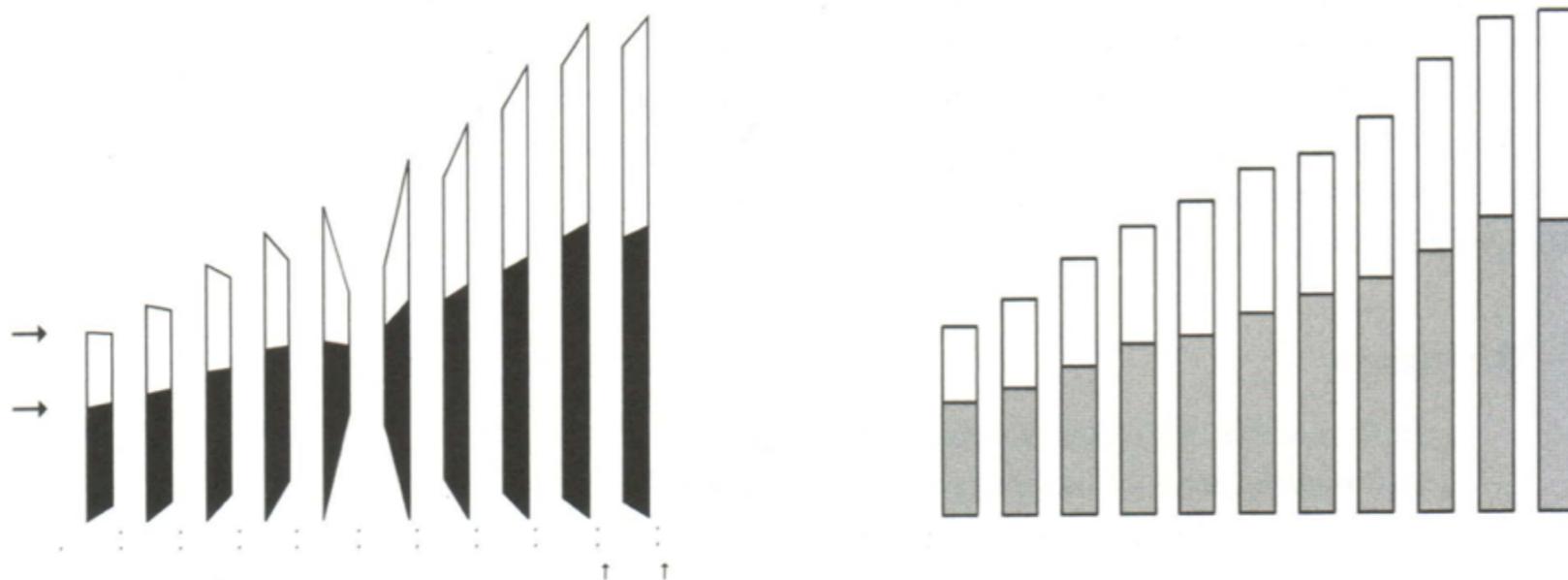
1. Barchart baseline fail



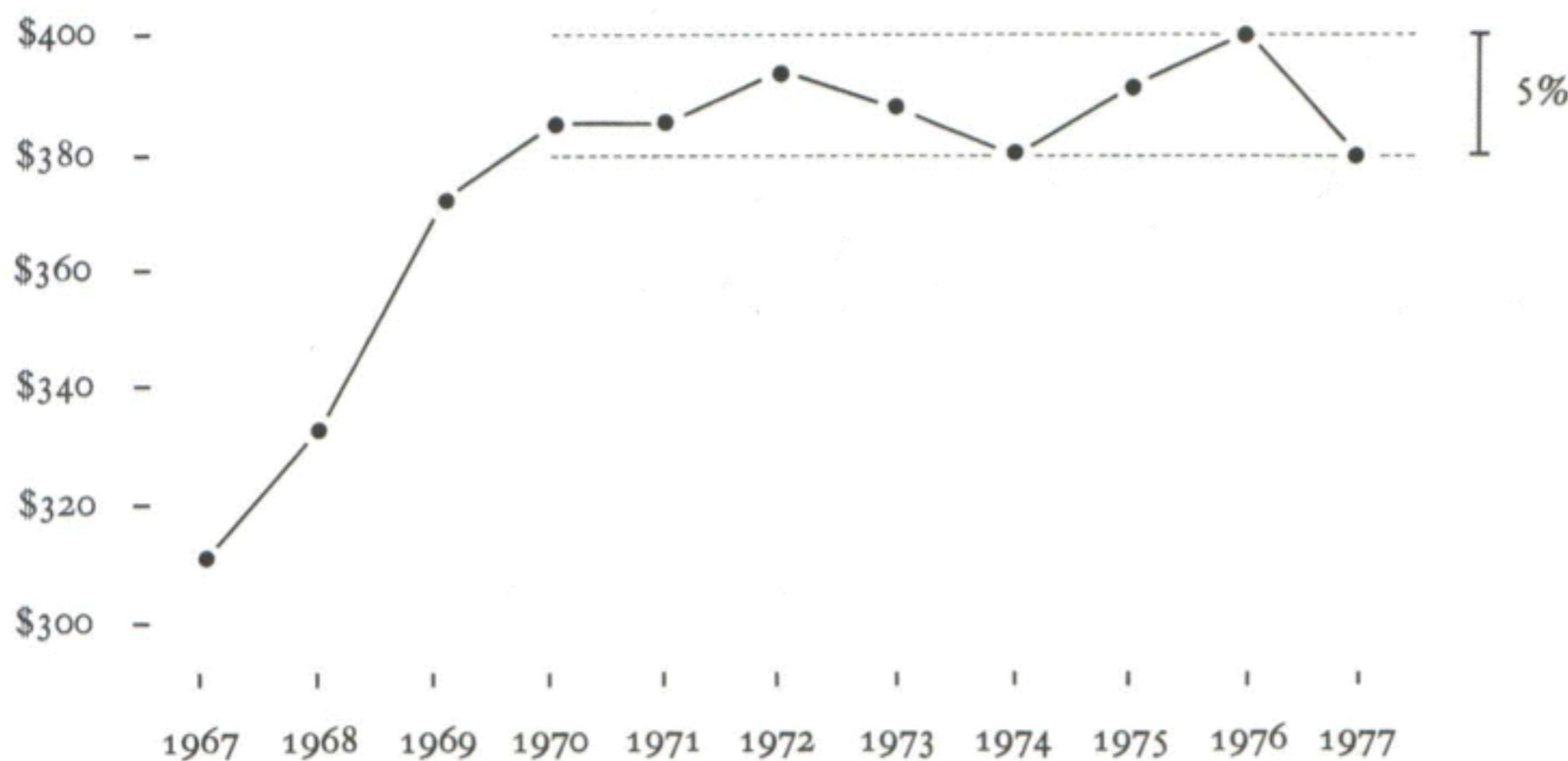
2. Perspective and measurement fail







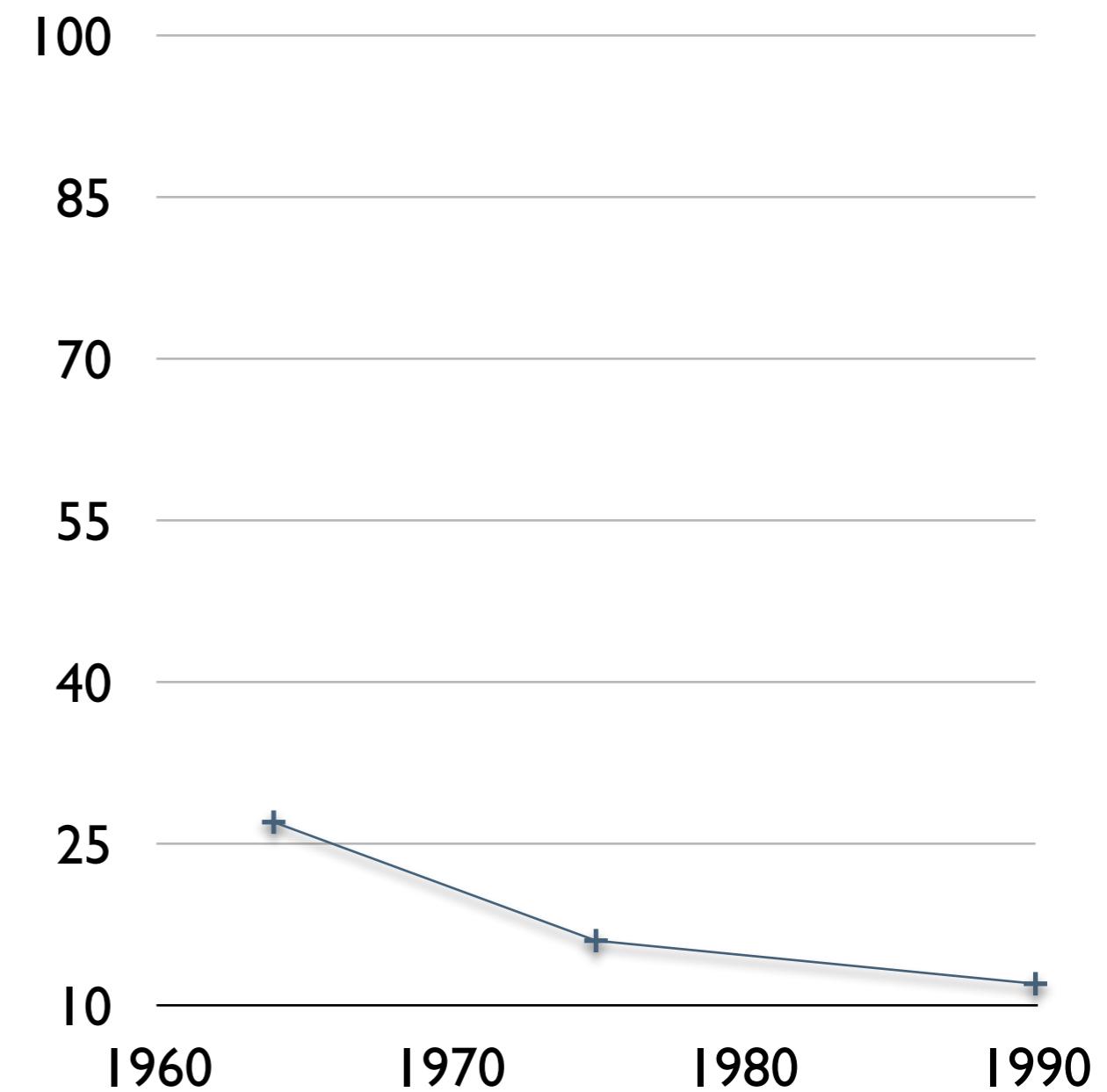
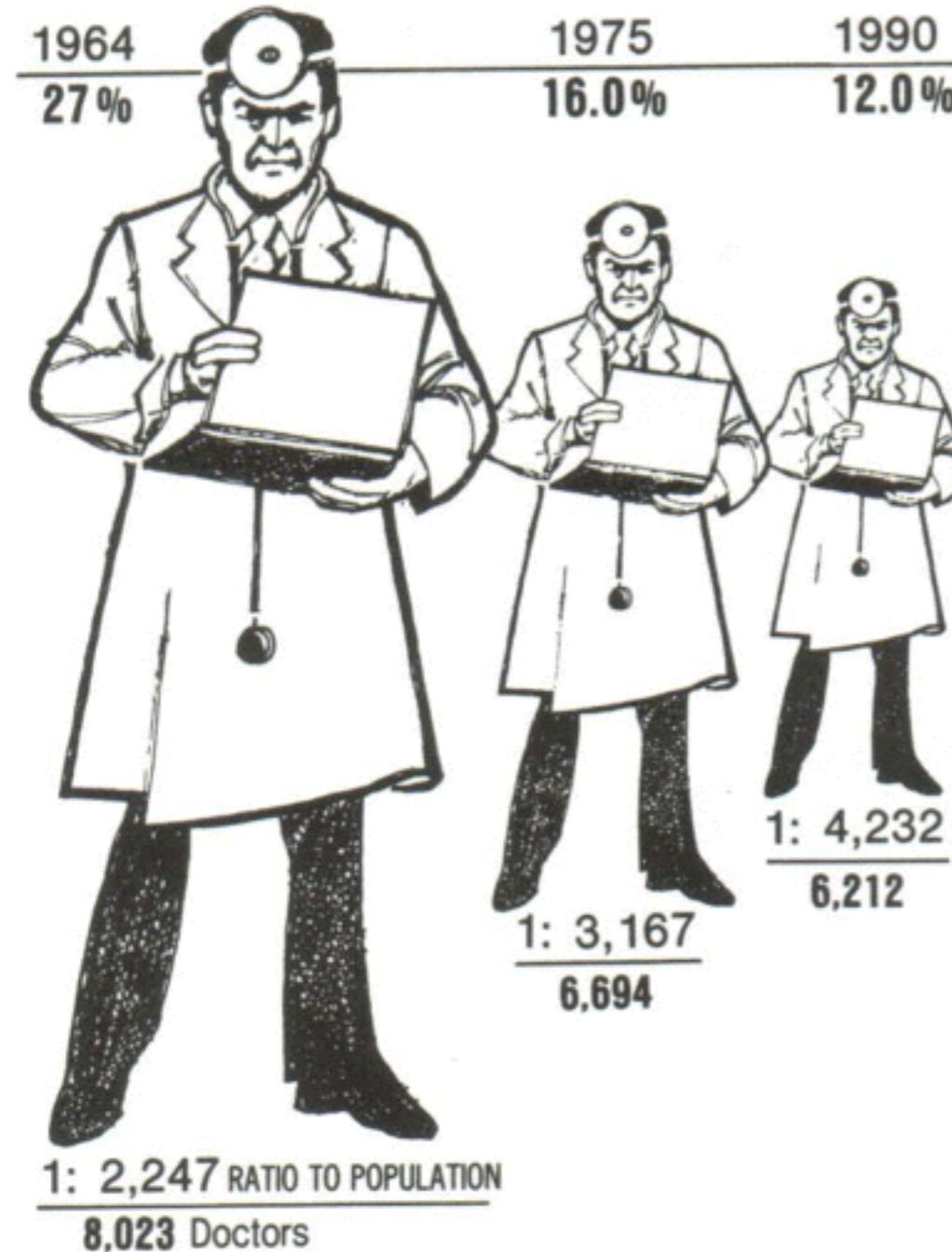
Per capita
budget expenditures,
in constant dollars



2. “Huge differences” fail

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice



using area (2 dimensions)



POLICY
PRACTICE



2.3
5.8



3

7



1

3.7



.8

3.8

5



7

NON-DISCRIMINATION IN
EMPLOYMENT & OCCUPATION

ILO Convention III (average score)

using area to represent one dime

Quiz: How does this fail?

THE ISSUE OF TRUST

ACCENTS AND DISTRUST

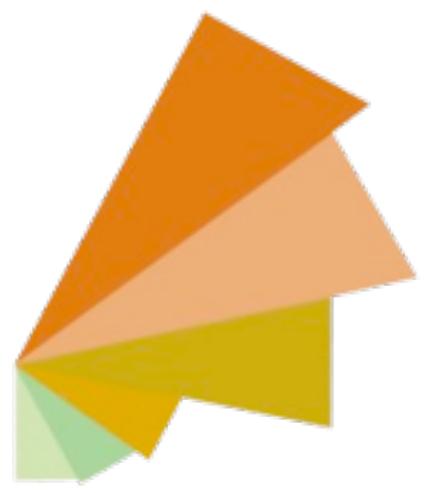
Another reason why accents affects customer service is the question of credibility. If I can not understand you, then I can not trust you.

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.

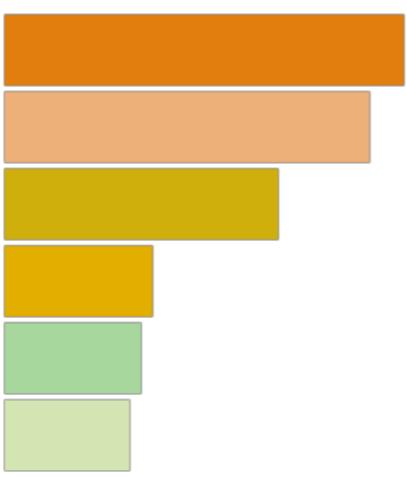
The results showed that the heavier the accent the less trust worthy the person became.

- ▲ NATIVE ACCENT
- ▲ MILD ACCENT
- ▲ HEAVY ACCENT





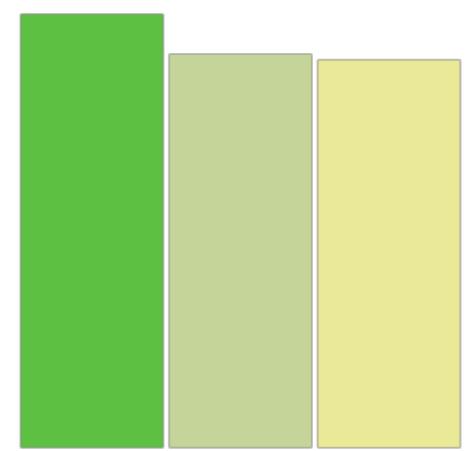
Distortion factor: 2.5



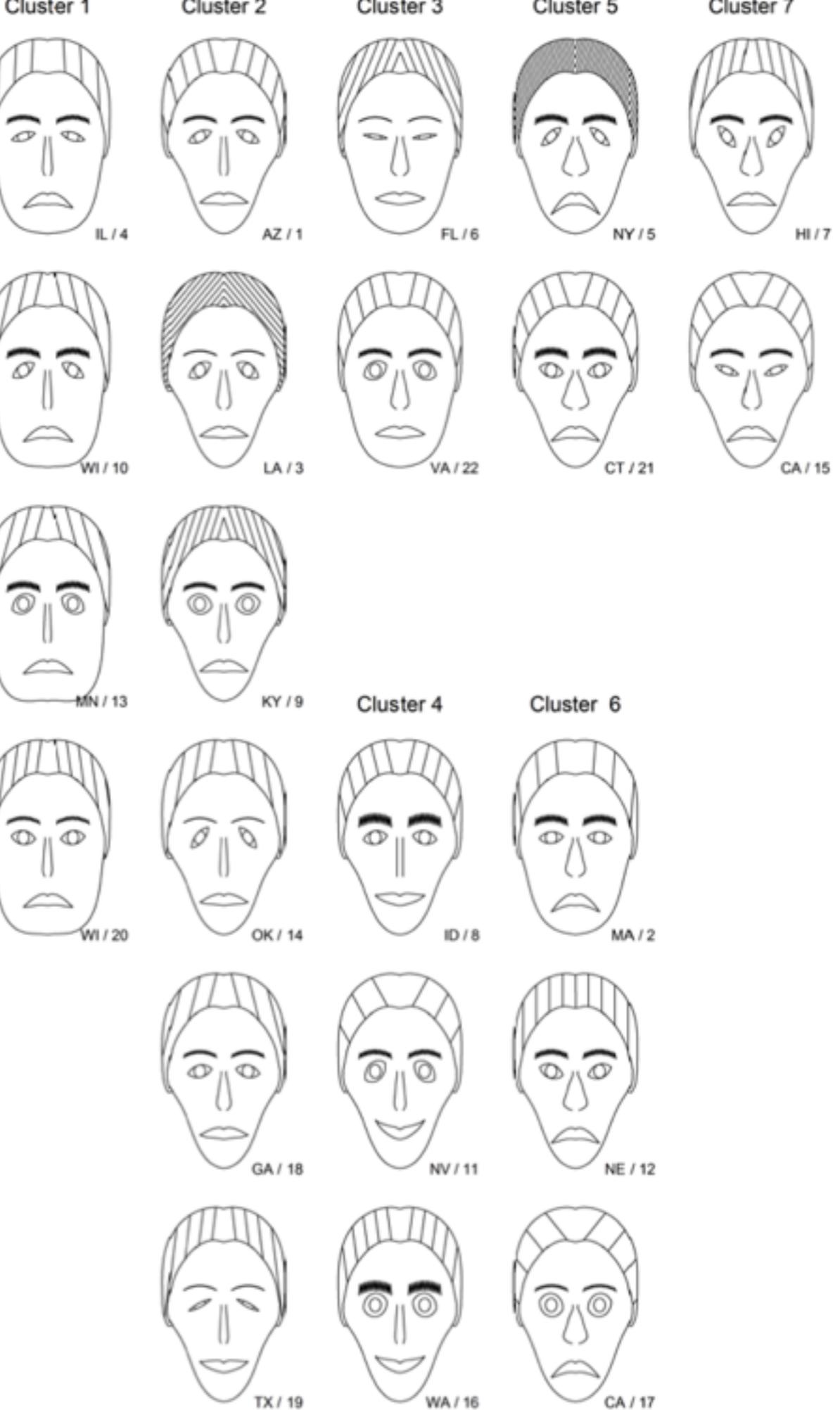
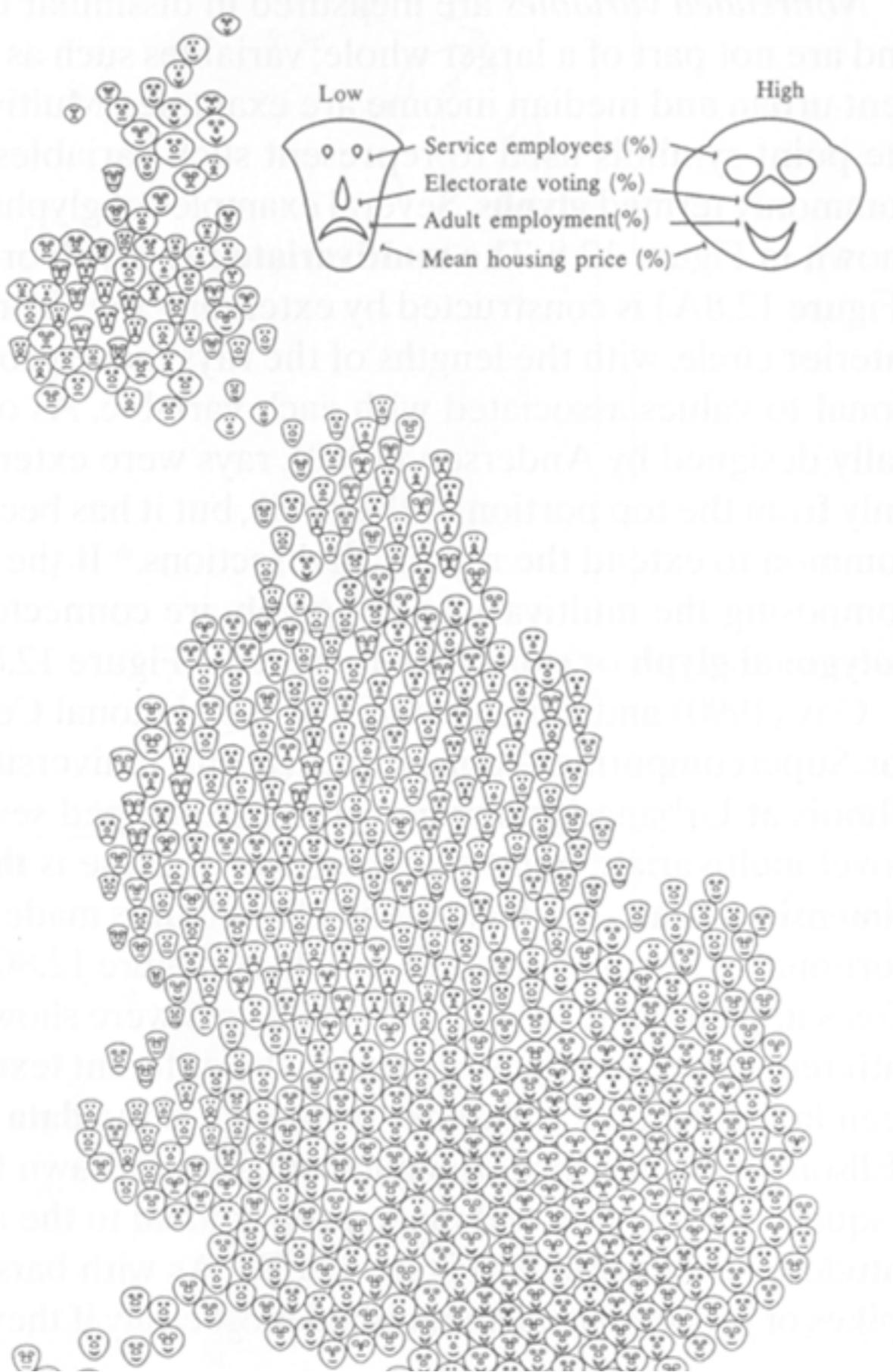
True data



Distortion factor: 5.0

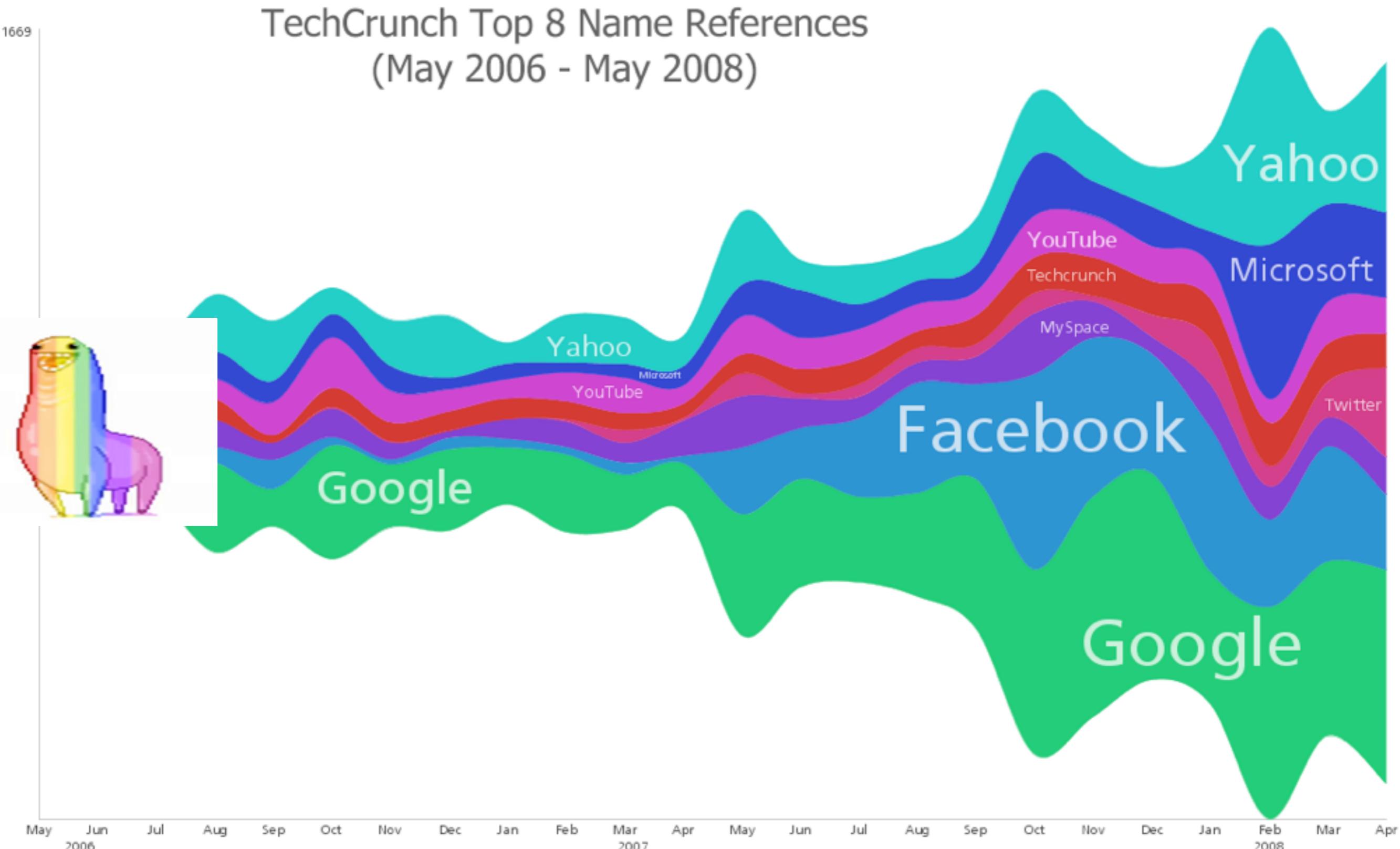


True data



Chernoff Faces

“streamgraphs”: double-stacked areas of horror



“abandon all hope ye who vieweth”