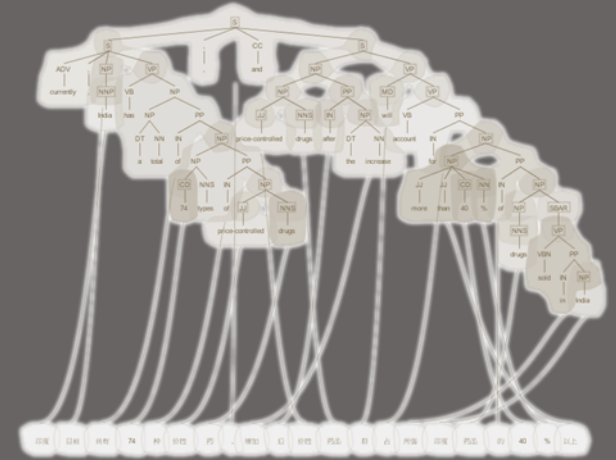
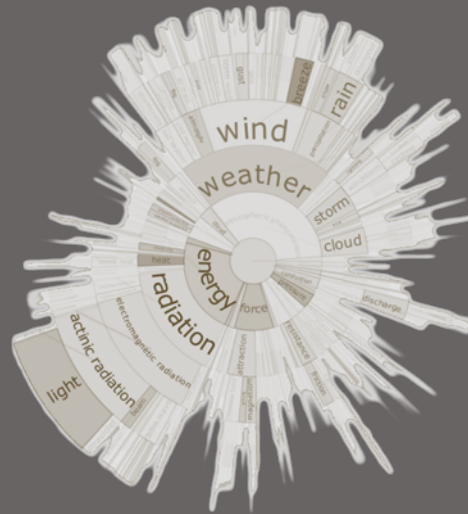
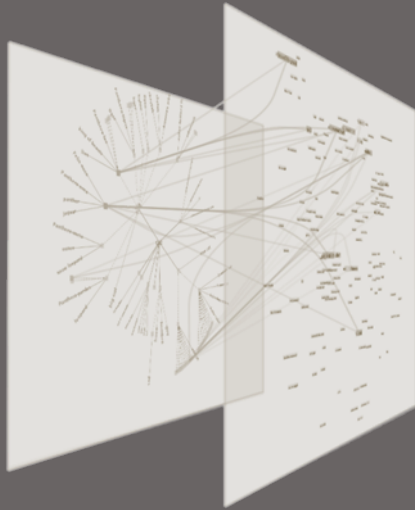




Interactive Visualization for Computational Linguistics

ESSLII 2009

Outline

2

DAY #1

- Introduction (G. Penn)
- Externalization (S. Carpendale)
- Exercise #1
- Break
- Information Visualization Basics (S. Carpendale)

DAY#2

- Review of Linguistic Visualizations (G. Penn)
- Critique #1
- Break
- Excellence in information graphics (S. Carpendale)

Outline

3

DAY #3

- Toolkits (G. Penn)
Break
- Critique #2
- ManyEyes

DAY #4

- Interaction (S. Carpendale)
- Critique #III
Break
- Excellence(S. Carpendale)

Outline

4

DAY #5

- Future directions (G. Penn)

Break

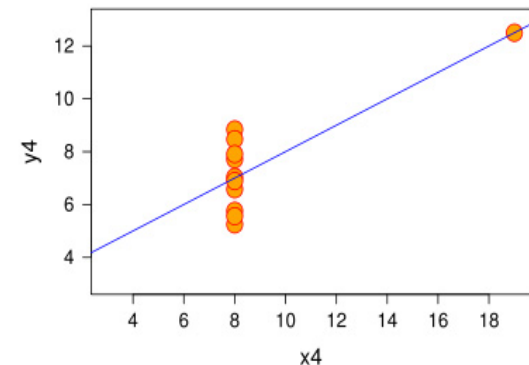
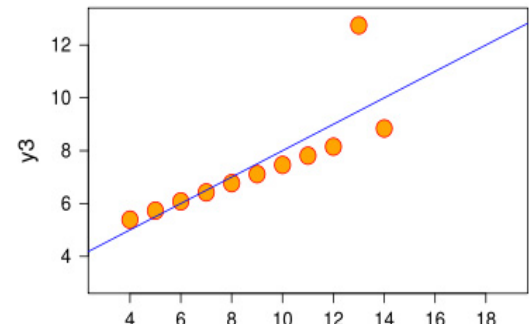
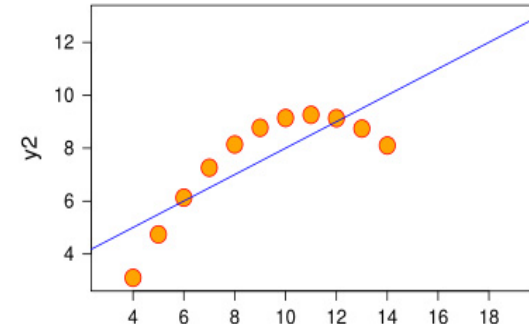
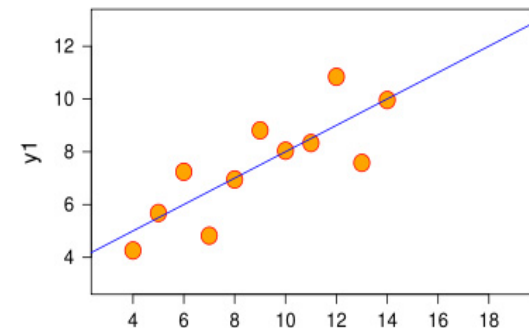
- Critique #IV
- Lessons learned (G. Penn and S. Carpendale)

The Power of Visualization

Anscombe's Quartet

6

I		II		III		IV	
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



7



London Cholera Deaths

8



John Snow (1855) On the Mode of Communication of Cholera.

Externalization

Externalization

10

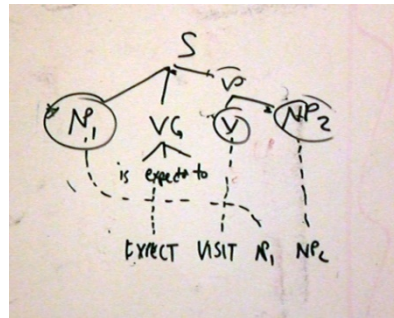
- External cognition is the interaction between internal and external representations when performing cognitive tasks.
- Computational offloading is the extent to which external representations can reduce the amount of cognitive effort to solve a problem.

Complexity Brings Externalization

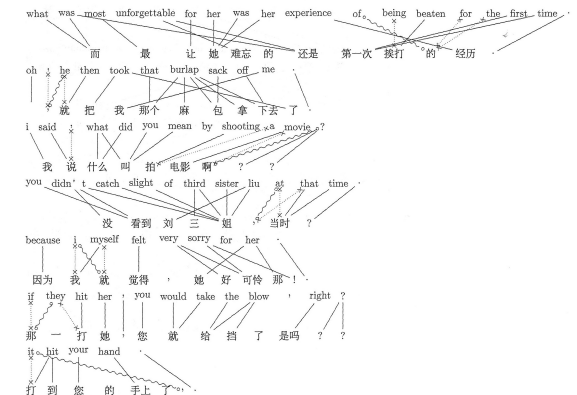
11

Simple:
analyze
internally

Complex:
sketch to aid
understanding



Very complex or
very large: Cannot
practically sketch
manually



Augmented Cognition

12

“The power of the unaided mind is highly overrated. Without external aids, memory thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities.”

Information Visualization

13

“Information visualization is the use of computer-supported interactive visual representations of abstract data to amplify cognition.”

... a form of external cognitive aid.

Functions of Visualization

14

- Recording information

 嬪 Tables, blueprints, satellite images

- Processing information

 嬪 needs feedback and interaction

- Presenting information

 嬪 share, collaborate, revise

 嬪 for oneself, for one's peers and to teach

- Seeing the unseen

Creating a Visualization

15

1. Understand a **system of related information and tasks**.
2. Create a **mapping** from the **data** (digital representation) to a **visual representation**.
3. **Present** this visual representation on the computer screen.
4. Provide **methods of interacting** with this **visual representation** that can include methods for varying the **presentation** and methods for varying the **representation**.
5. **Verify** the usefulness of the **representation**, the way it is **presented** and/or and its **interaction methods**.

Exercise #1

Document	0	1	2	3	4	5	6	7	8	9
1st pers	1.37	0.89	4.86	0.88	10.23	15.23	2.59	16.98	7.42	4.01
2nd pers	2.73	9.83	0.88	0.00	0.89	0.00	2.59	19.40	0.44	0.45
3rd pers	34.64	16.09	13.26	22.81	26.68	18.81	22.45	59.42	20.07	23.17
past verbs	50.14	34.87	45.96	64.04	63.58	72.55	37.13	77.61	63.26	76.65
saying verbs	3.19	0.00	2.21	7.02	7.11	0.90	1.30	8.49	6.11	7.58
nouns	205.10	216.36	269.11	289.04	306.36	303.18	256.04	200.08	300.61	294.12
adverbs	47.40	46.94	35.35	36.84	30.68	26.42	41.88	50.12	31.85	35.65
that	16.86	5.36	4.42	4.82	8.89	4.93	11.23	6.87	9.60	9.36
hifalutin	40.11	21.01	71.59	40.35	46.69	52.84	32.82	7.68	38.39	55.26
sent len	22.85	22.15	26.01	21.31	23.43	25.67	19.79	9.86	25.19	24.13
genre	official	feature	official	reportage	reportage	reportage	official	fiction	reportage	reportage

Key:	1st/2nd/3rd pers, past/saying verbs, nouns, adverbs, that	frequency counts per 1K tokens
	hifalutin	frequency counts per 1K tokens of nouns and verbs with at least 10 orthographic characters (at least 12 for VBGs)
	sent len	mean length of sentences in tokens
	genre	one of {reportage, feature (belles-lettres), official, fiction}

BREAK

19

Information Visualization Basics

Visual Variables

References

21

- Slides in this section are based on:

Semiology of Graphics, Jacques Bertin, 1983
translation of *Semiologie Graphique*, 1967

- With some extensions

Representation

- A representation is
 - **a formal system or mapping by which the information can be specified** (D. Marr)
 - **a sign system in that it stands for something other than its self.**
- for example: the number thirty-four
 - decimal: 34,
 - binary: 100010,
 - roman: XXXIV
- different representations reveal different aspects of the information
 - decimal: counting & information about powers of 10,
 - binary: counting & information about powers of 2,
 - roman: counting & adding and subtracting
- presentation
 - how the representation is placed or organized on the screen

34, **34**, **34**

Representations

- Solving a problem simply means representing it so as to make the solution transparent ... (*Simon, 1981*)
- Good representations
 - 嬪 allow people to *find* relevant information
 - information may be present but hard to find
 - 嬪 allow people to *compute* desired conclusions
 - computations may be difficult or “for free” depending on representations

Creating Visual Representations

24

- To communicate with words we first need to know phonemes, the letters and how they combine to create words
 - 嬪 note that phonemes and letters are “meaningless” in themselves
- are there corresponding visual units?
 - 嬪 there is still considerable debate on this subject
- in the meantime, we will look at the practical approach of Jacques Bertin on how we can create visual representations that can be understood.

Disclaimer

25

- **Bertin considers:**

- 嬪 printable, on white paper,

- 嬪 visible at a glance

- 嬪 reading distance of book or atlas

- 嬪 normal and constant lighting

- 嬪 readily available graphic means

- Good as a guideline, we need to consider context and medium of visualization use.

Where do we start?

26

- With marks!
 - 嬪 for us, pixels
- *Visual Variables*: how can we vary marks?
 - 嬪 by where we place them
 - 嬪 by how we place them
 - 嬪 by their visual characteristics

The Plane

27

- Points
- Lines
- Areas

Points

28

- “A point represents a location on the plane that has no theoretical length or area. This signification is independent of the size and character of the mark which renders it visible.”
- a location
- marks that indicate points can vary in all visual variables

Lines

29

- “A line signifies a phenomenon on the plane which has measurable length but no area. This signification is independent of the width and characteristics of the mark which renders it visible.”
- a boundary, a route, a connection

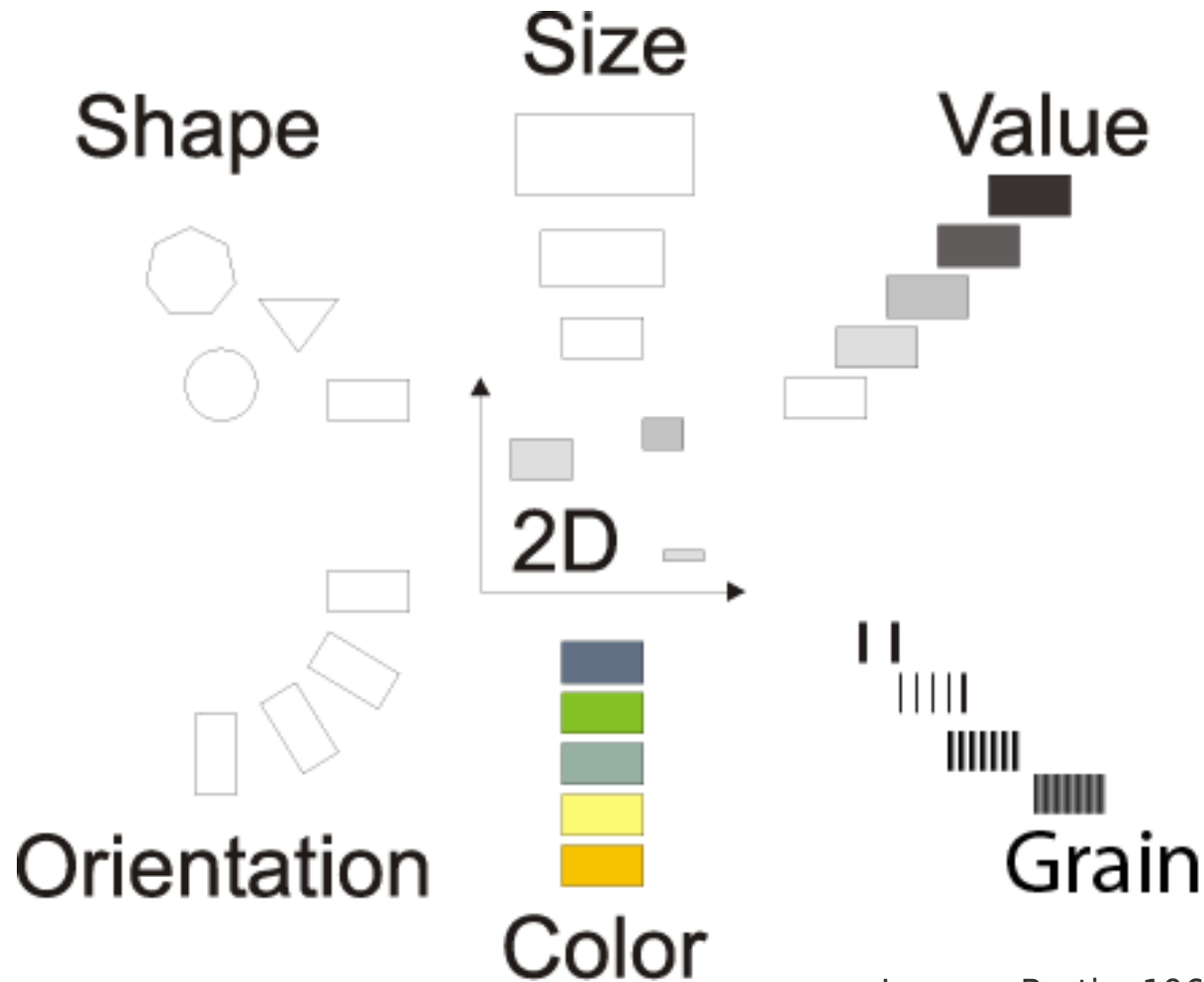
Areas

30

- “An area signifies something on the plane that has measurable size. This signification applies to the entire area covered by the visible mark.”
- an area can change in position but not in size, shape or orientation without making the area itself have a different meaning

Visual Variables

31



Jacques Bertin, 1967; translated 1983

Additional Variables for Computers

32

- **motion**

嬪 direction, acceleration, speed, frequency, onset, 'personality'

- **saturation**

嬪 colour as Bertin uses it largely refers to hue, other readily available colour channels (*i.e.* saturation)

- **flicker**

嬪 frequency, rhythm, appearance

- **depth? 'quasi' 3D**

嬪 depth, occlusion, aerial perspective, binocular disparity

- **illumination**

- **transparency**

Characteristics of Visual Variables

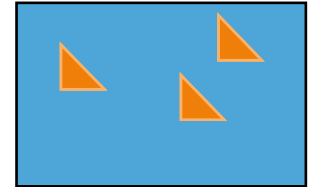
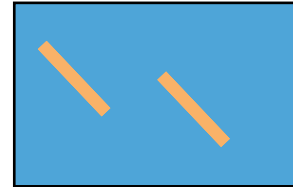
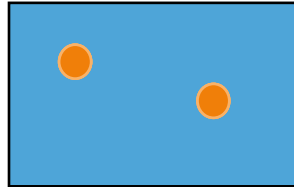
33

- **Selective:**
Is a change in this variable enough to allow us to select it from a group?
- **Associative:**
Is a change in this variable enough to allow us to perceive them as a group?
- **Quantitative:**
Is there a numerical reading obtainable from changes in this variable?
- **Order:**
Are changes in this variable perceived as ordered?
- **Length:**
Across how many changes in this variable are distinctions possible?

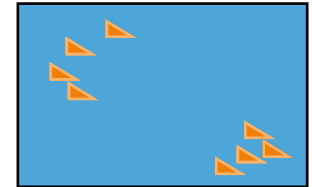
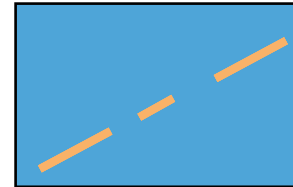
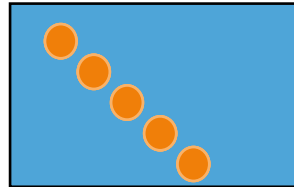
Visual Variable: Position

34

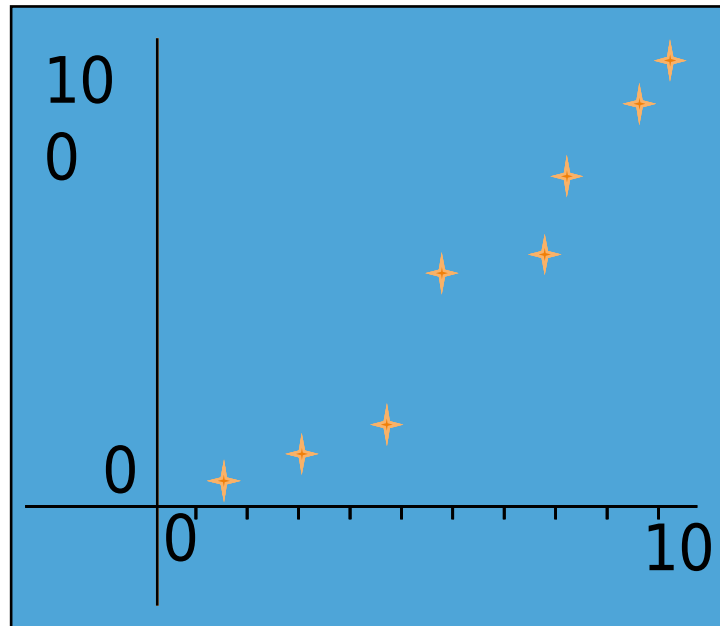
✓ ☐ selective



✓ ☐ associative



✓ ☐ quantitative



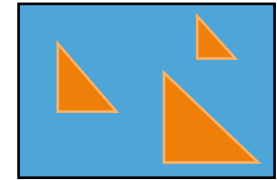
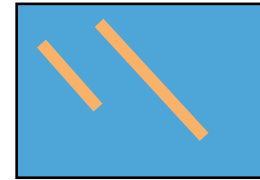
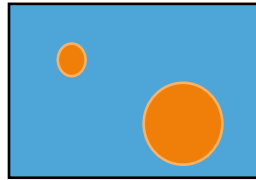
✓ ☐ order

✓ ☐ length

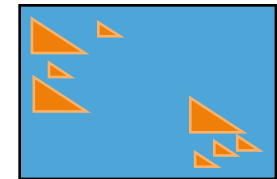
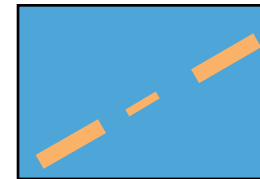
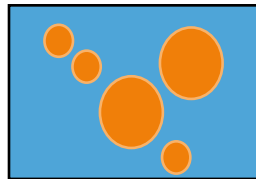
Visual Variable: Size

35

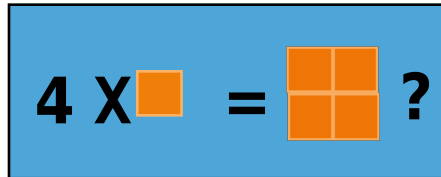
✓ □ selective



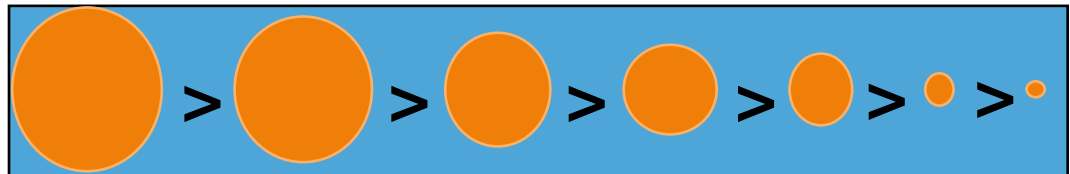
✓ □ associative



✗ □ quantitative



✓ □ order



✓ □ length

續 theoretically infinite but practically limited

續 association and selection ~ 5 and distinction ~ 20

Size

36



points



lines

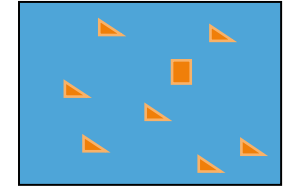
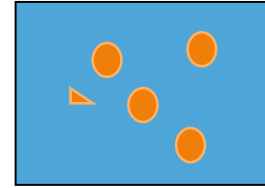
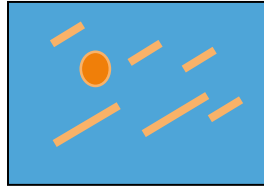


areas

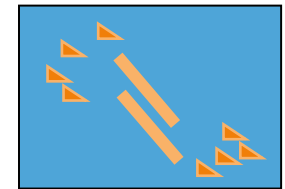
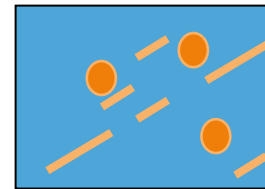
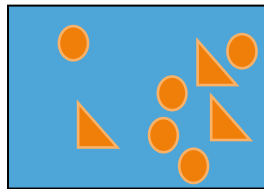
Visual Variable: Shape

37

 ☐ selective



 ☐ associative



 ☐ quantitative



 ☐ order



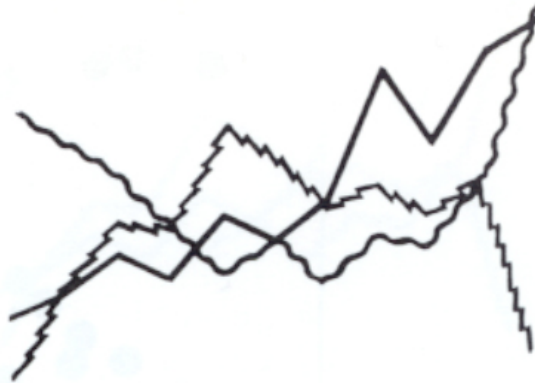
 ☐ length
  infinite

Shape

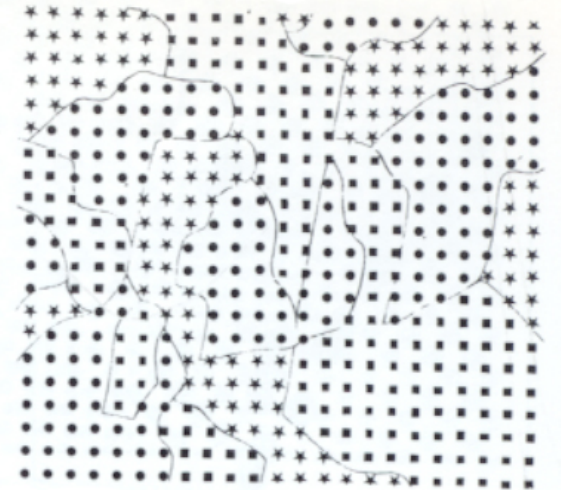
38



points



lines



areas

an aside (perception)

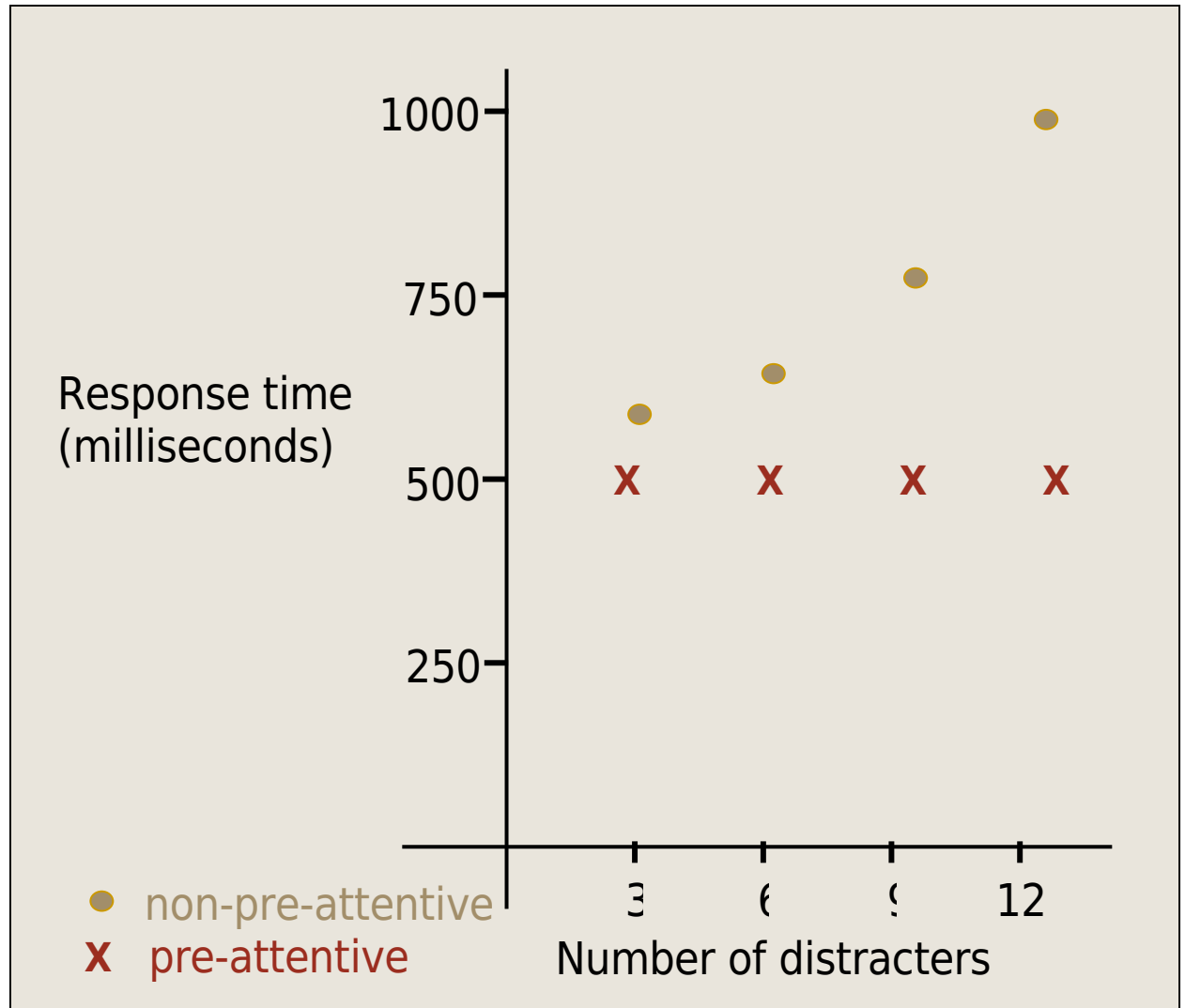
Pre-attentive processing

40

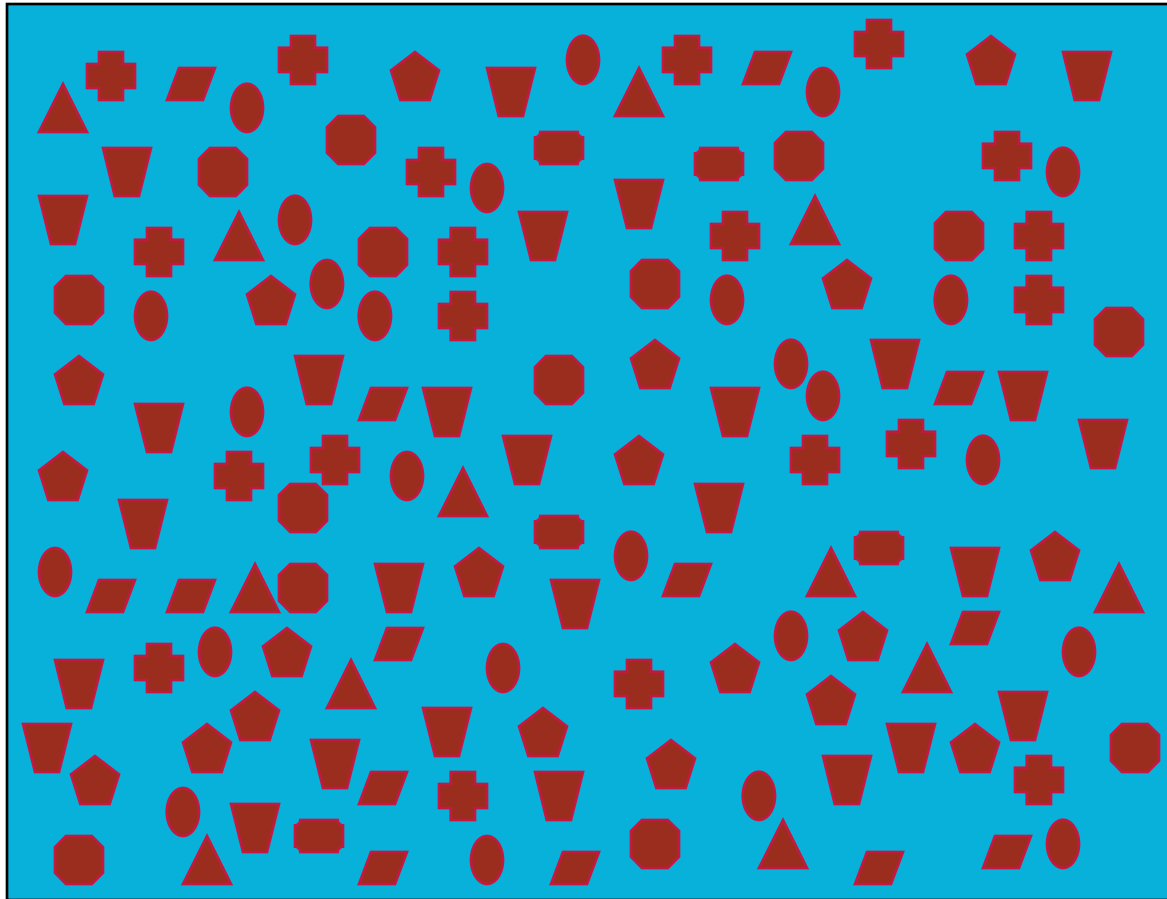
2358945739756860796752453512346534624356245762457
2456134523523523523523524351345324716498762987460
3295872358276533637872138764298769876364098721696
5329624139237462163987639871236597124593874638746
9887126498172649872165971523972356987129721653978
2164098712464783467218987639450897764398217346946
4964392764309872634287469864987597152397123976490
8714698764987243698127346987461435895321456865437

2358945739756860796752453512346534624356245762457
2456134523523523523523524351345324716498762987460
3295872358276533637872138764298769876364098721696
5329624139237462163987639871236597124593874638746
9887126498172649872165971523972356987129721653978
2164098712464783467218987639450897764398217346946
4964392764309872634287469864987597152397123976490
8714698764987243698127346987461435895321456865437

Typical Results



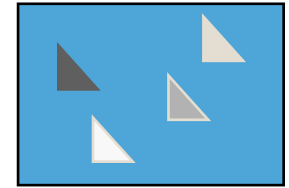
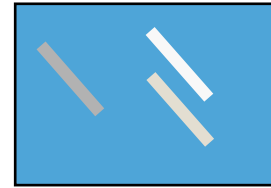
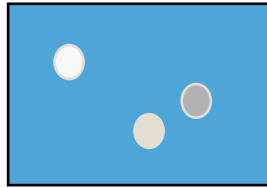
Shape



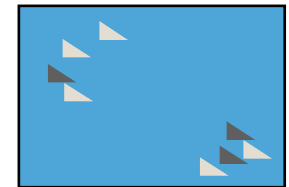
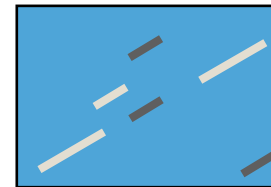
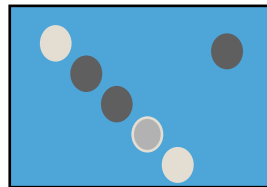
Visual Variable: Value

43

✓ □ selective

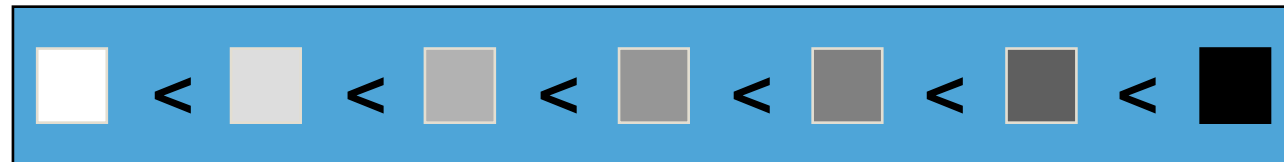


✓ □ associative



✗ □ quantitative

✓ □ order



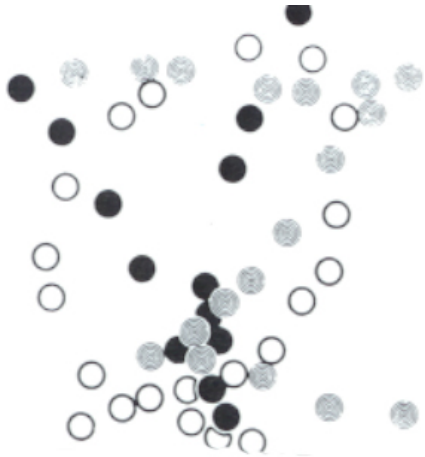
✓ □ length

𐄂 theoretically infinite but practically limited

𐄂 association and selection ~ < 7 and distinction ~ 10

Value

44



points



lines

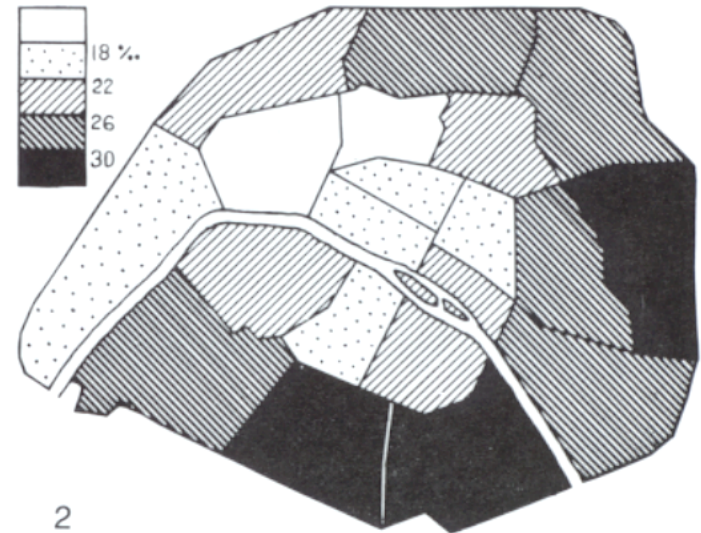
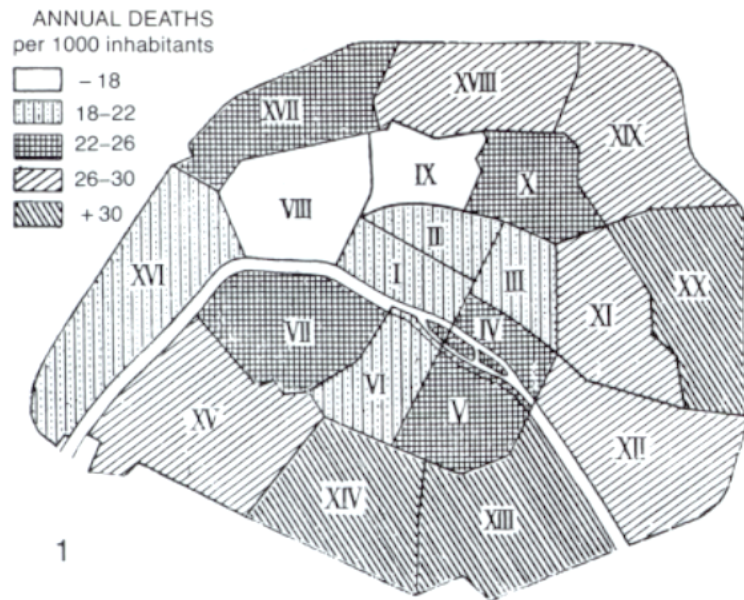


areas

Value

45

- Ordered, cannot be reordered



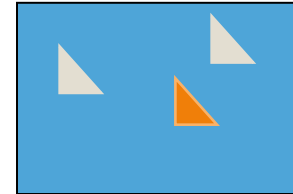
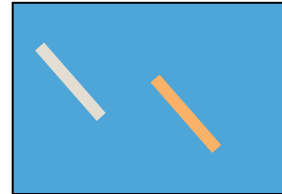
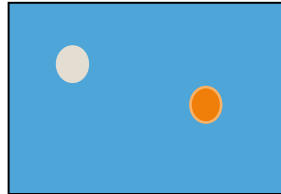
annual deaths per 1000 inhabitants, Paris

Visual Variable: Colour

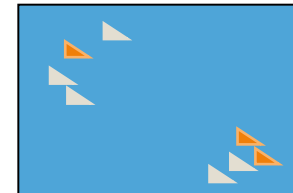
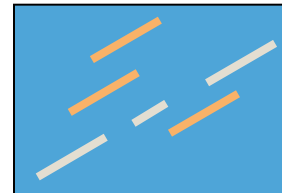
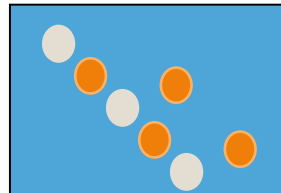
46



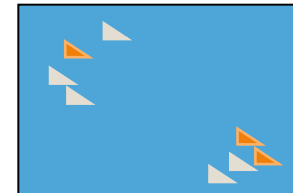
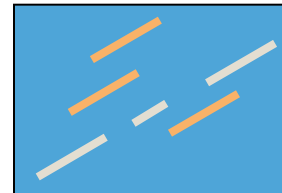
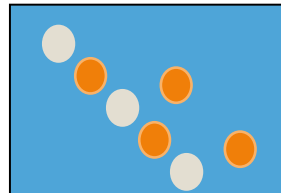
□ selective



□ associative



□ quantitative



□ order

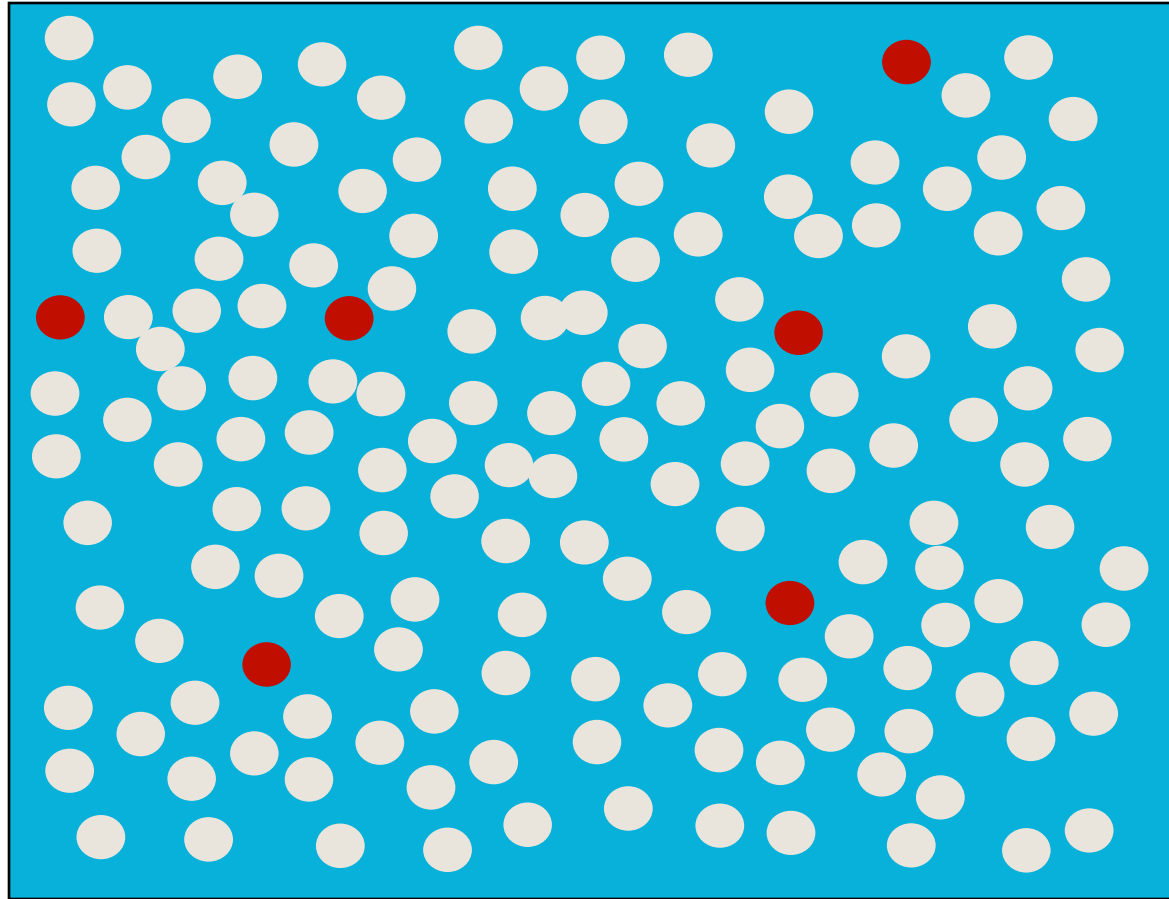


□ length

續 theoretically infinite but practically limited

續 association and selection $\sim < 7$ and distinction ~ 10

Colour



Colour

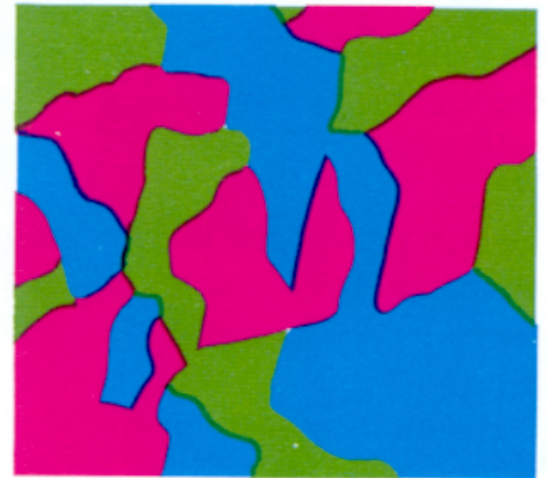
48



points



lines

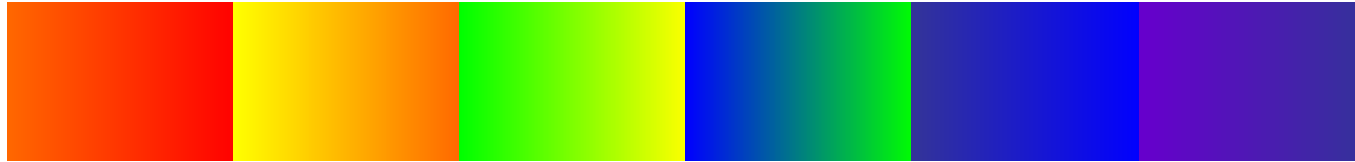


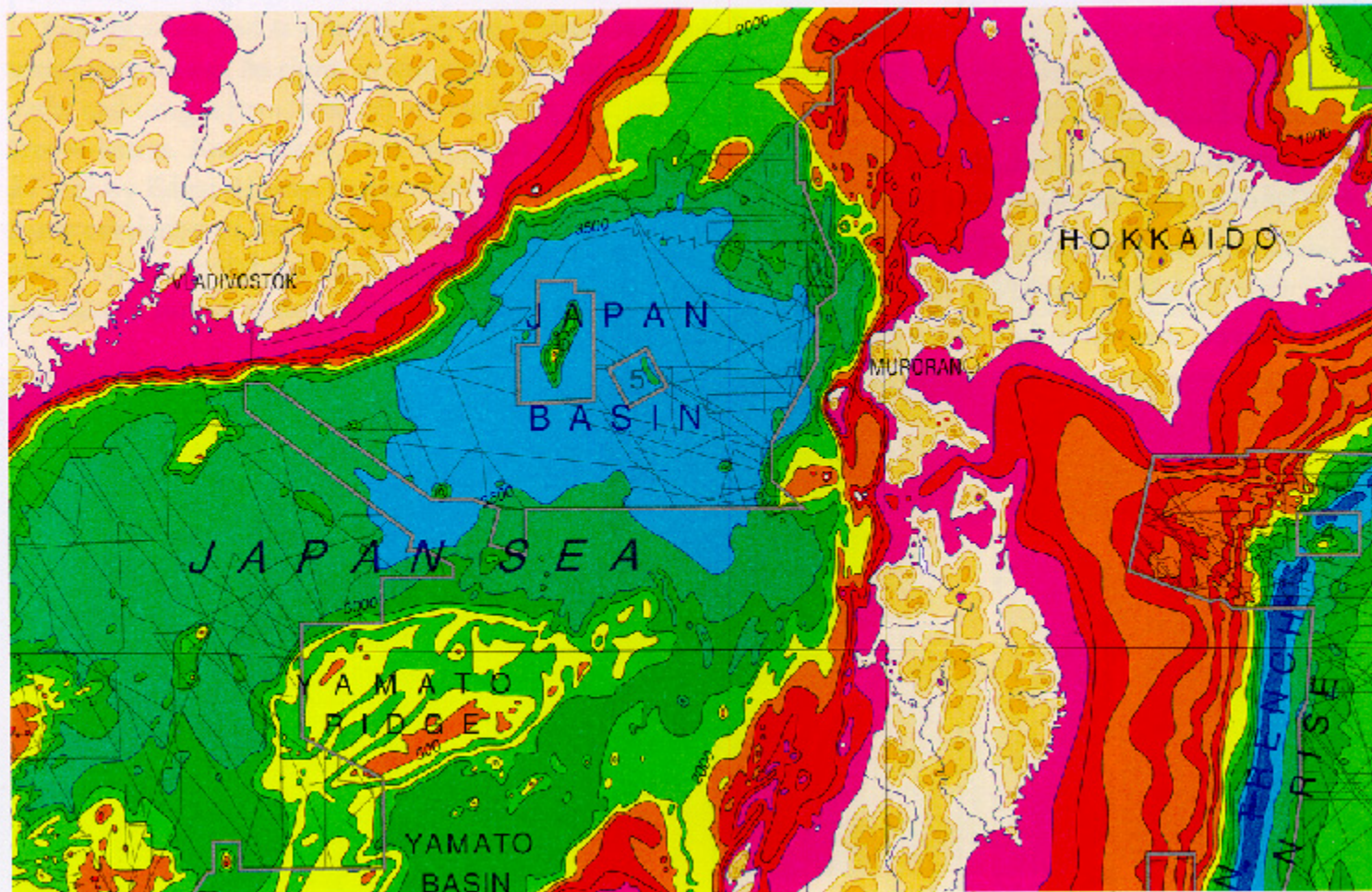
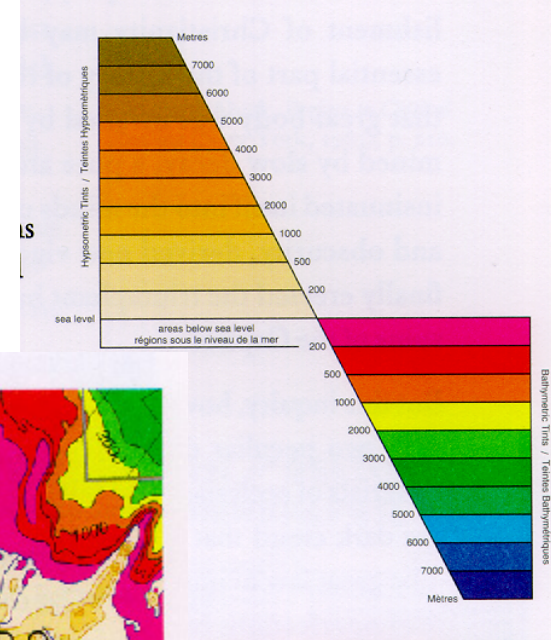
areas

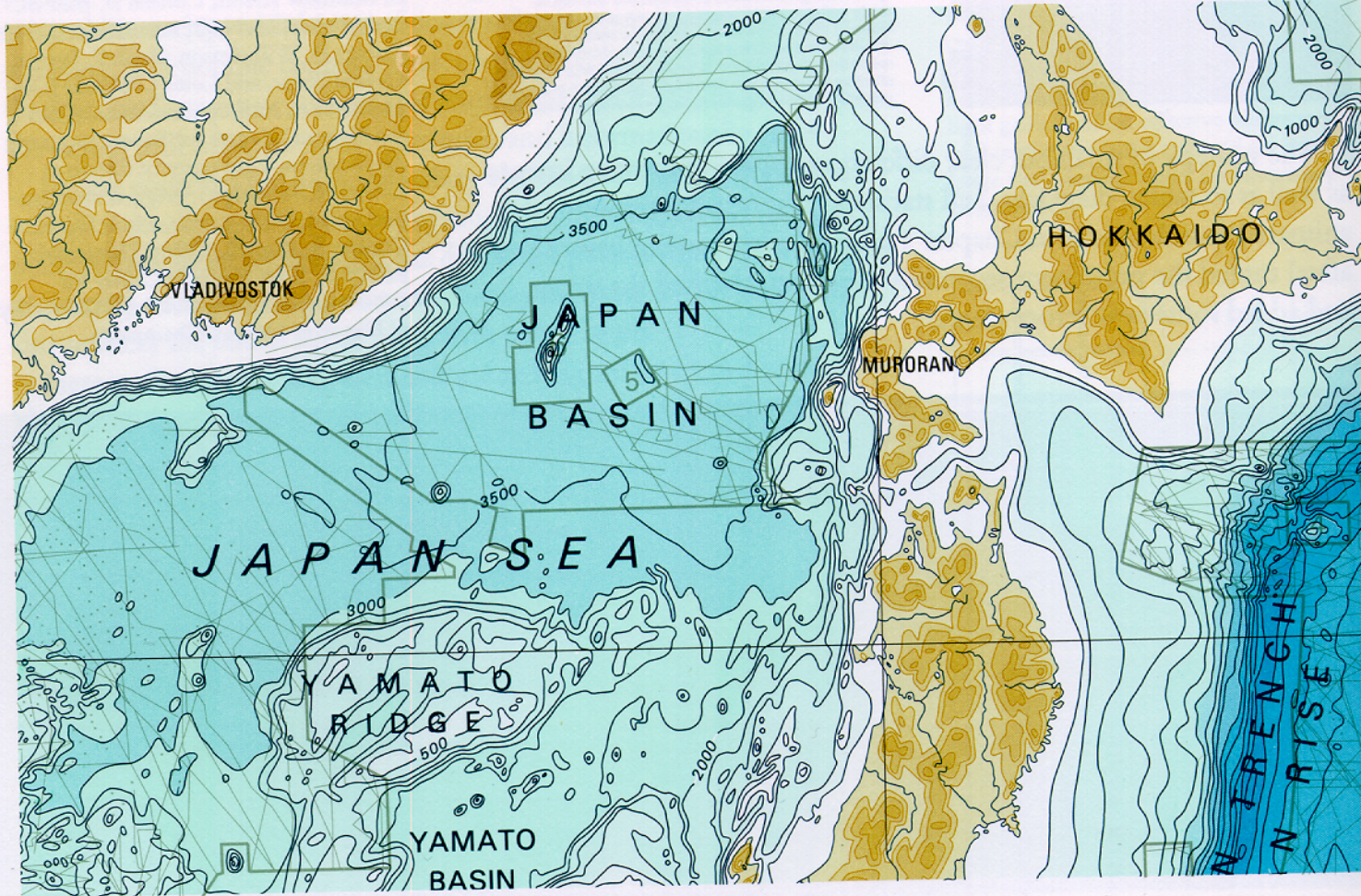
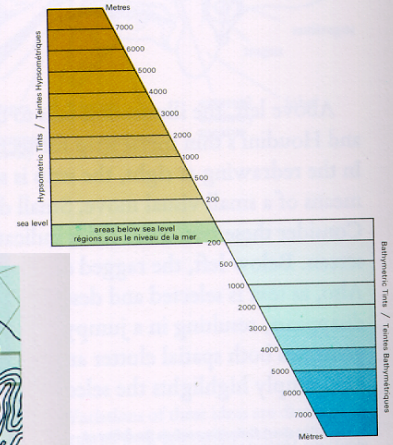
Colour Scales

49

- Common to use a rainbow scale... caution!

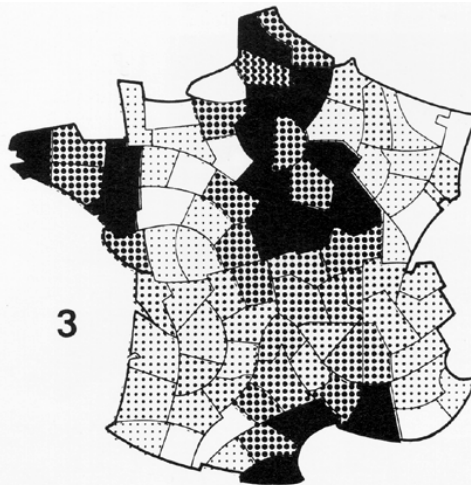




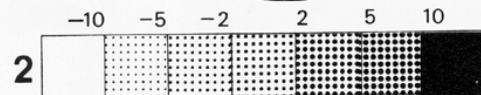




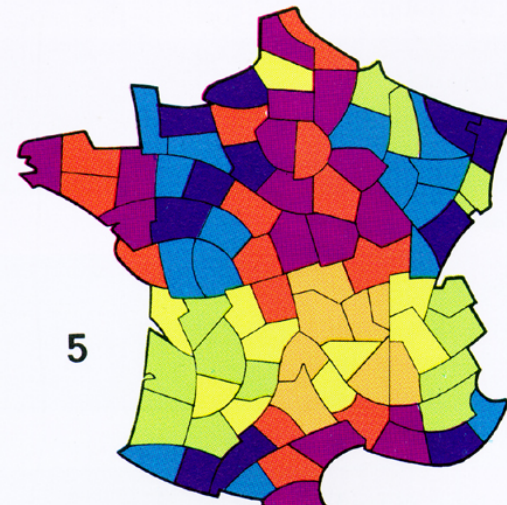
1



3



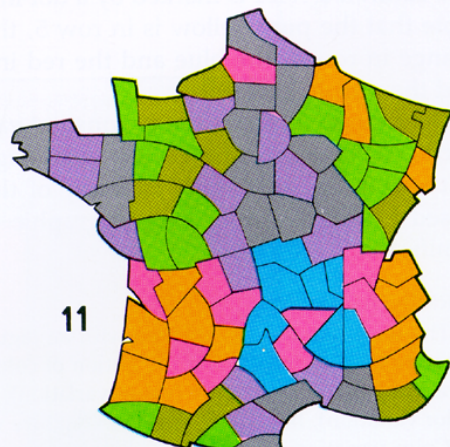
2



5



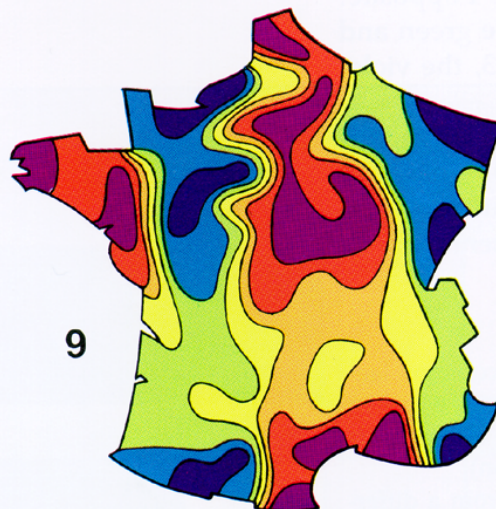
4



11



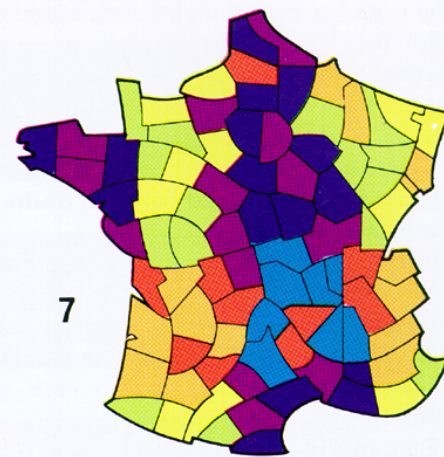
10



9



8



7



6

Controls Legibility

53



Colour Blindness

54

- VisCheck.com (Robert Dougherty and Alex Wade)
 - 孀 Web service to simulate colour blindness
- Biggest problems with red/green
- Varying value and hue can help



Deuteranope



Protanope



Tritanope

Colour Resources

55

- Maureen Stone's Resources

嬪 <http://www.stonesc.com>

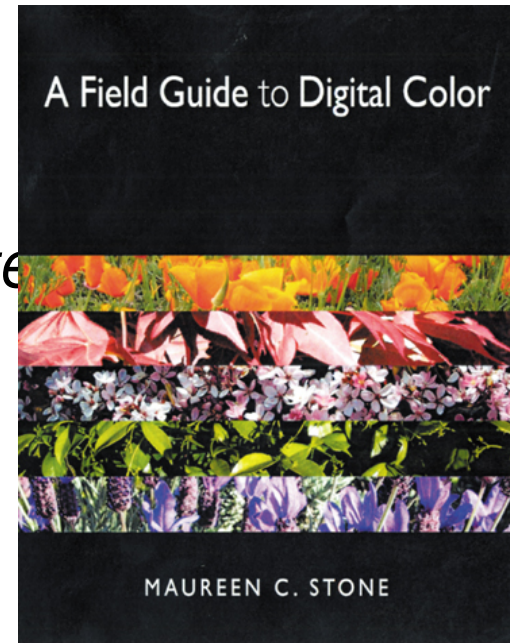
嬪 *A Field Guide to Digital Color* (A. K. Peter

- Cindy Brewer's *ColorBrewer*

嬪 <http://colorbrewer.org>

- ColourLovers Community Palette Sharing

嬪 <http://www.colourlovers.com>



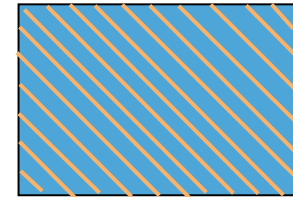
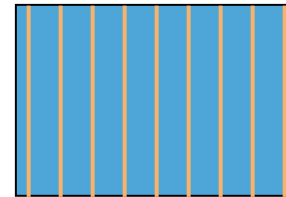
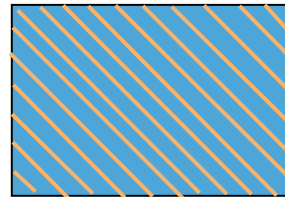
Visual Variable: Orientation

56

✓ □ selective

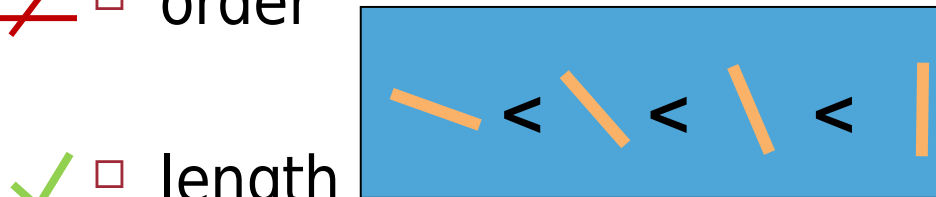


✓ □ associative

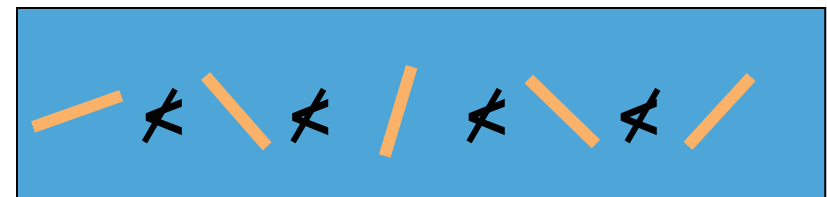


✗ □ quantitative

✗ □ order



?



✓ □ length

續 ~5 in 2D; ? in 3D

Orientation

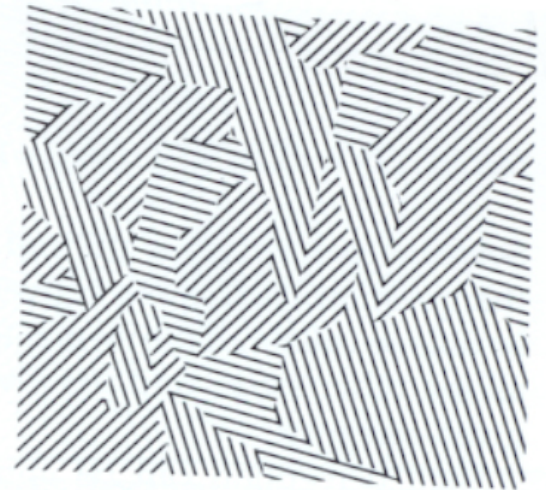
57



points



lines



areas

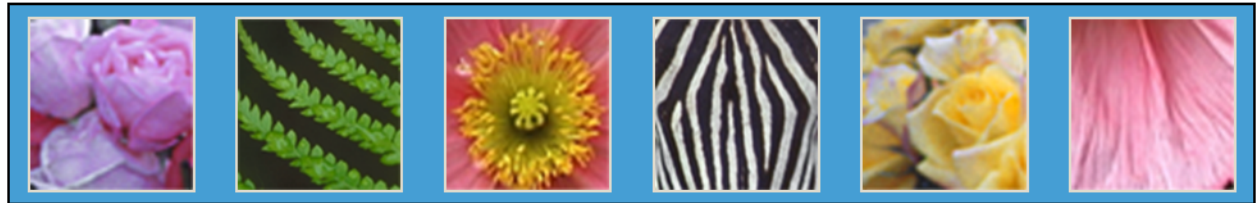
Visual Variable: Texture

58

✓ □ selective

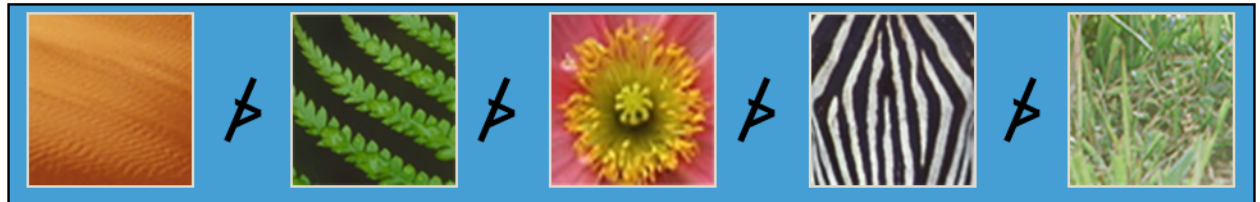


✓ □ associative



≠ □ quantitative

≠ □ order



✓ □ length

續 theoretically infinite

Texture

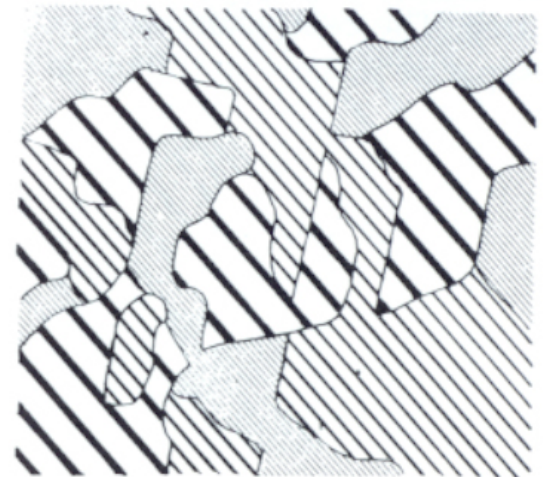
59



points



lines



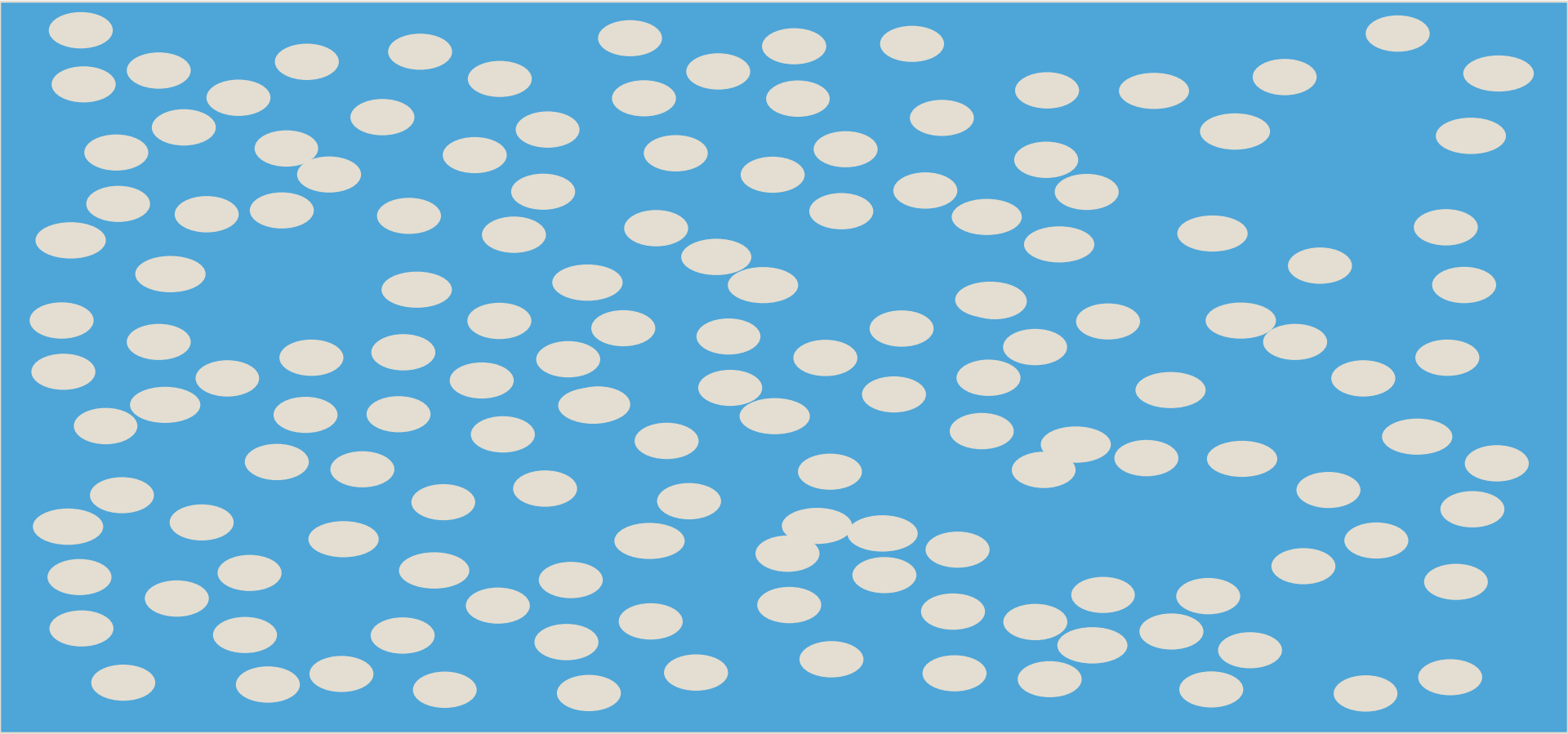
areas

Visual Variable: Motion

- ✓ □ selective
 - 嬪 motion is one of our most powerful attention grabbers
- ✓ □ associative
 - 嬪 moving in unison groups objects effectively
- ≠ □ quantitative
 - 嬪 subjective perception
- ≠ □ order
- ? □ length
 - 嬪 distinguishable types of motion?

Motion

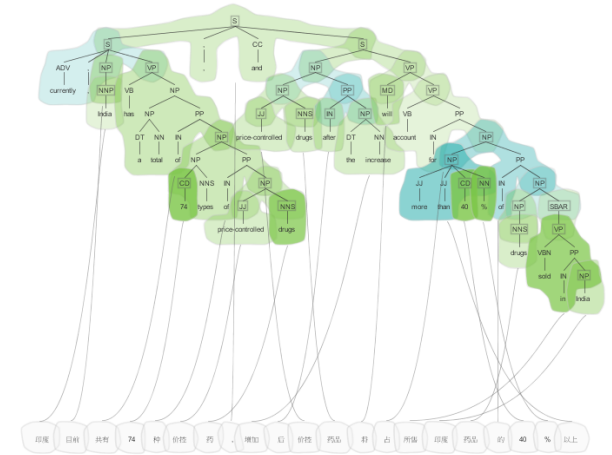
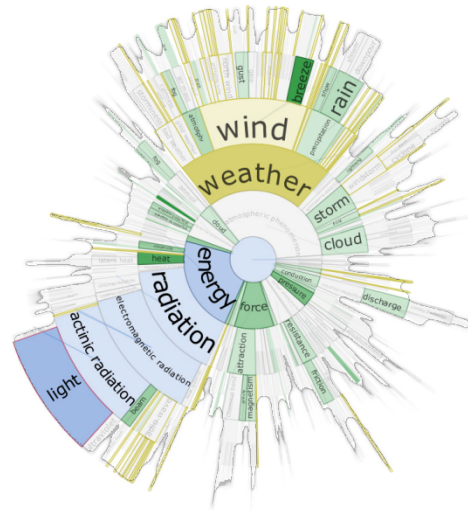
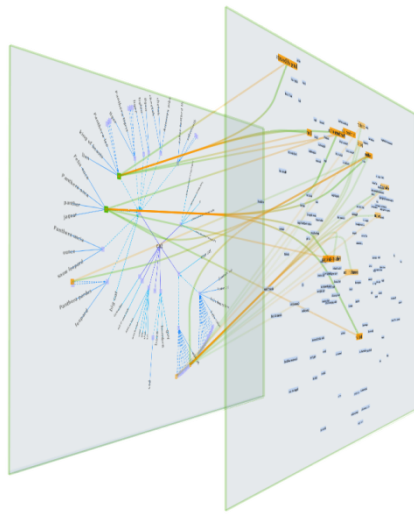
61



Visual Variables

62

Visual Variable	Selective	Associative	Quantitative	Order	Length
Position	Yes	Yes	Yes	Yes	Dependant on resolution
Size	Yes	Yes	Approximate	Yes	Association: 5; Distinction: 20
Shape	With Effort	With Effort	No	No	Infinite
Value	Yes	Yes	No	Yes	Association: 7; Distinction: 10
Hue	Yes	Yes	No	No	Association: 7; Distinction: 10
Orientation	Yes	Yes	No	No	4
Grain	Yes	Yes	No	No	5
Texture	Yes	Yes	No	No	Infinite
Motion	Yes	Yes	No	Yes	Unknown



*For tonight
make a drawing of the data set we looked at
earlier*

CHRISTOPHER COLLINS

CCOLLINS@CS.UTORONTO.CA

GERALD PENN

GPENN@CS.UTORONTO.CA

SHEELAGH CARPENDALE

SHEELAGH@UCALGARY.CA



Perception

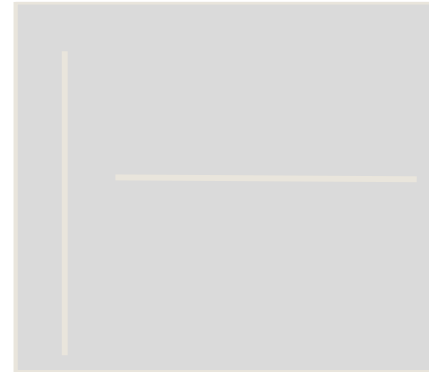
Visual Gestalt

proximity

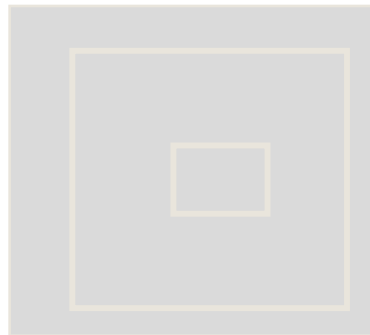


similarity

continuity



closure



figure/ground



symmetry

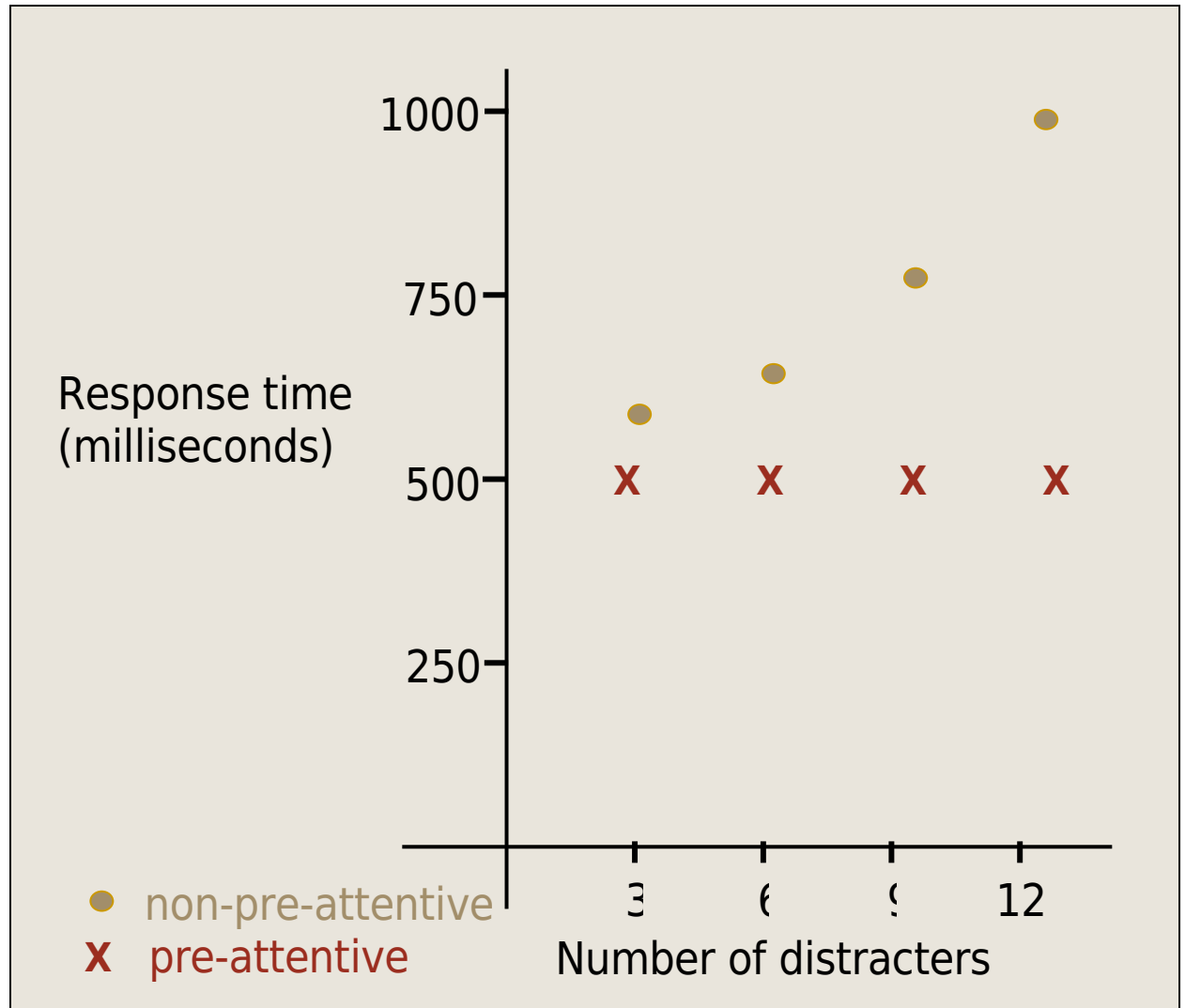
Pre-attentive processing

66

2358945739756860796752453512346534624356245762457
2456134523523523523523524351345324716498762987460
3295872358276533637872138764298769876364098721696
5329624139237462163987639871236597124593874638746
9887126498172649872165971523972356987129721653978
2164098712464783467218987639450897764398217346946
4964392764309872634287469864987597152397123976490
8714698764987243698127346987461435895321456865437

2358945739756860796752453512346534624356245762457
2456134523523523523523524351345324716498762987460
3295872358276533637872138764298769876364098721696
5329624139237462163987639871236597124593874638746
9887126498172649872165971523972356987129721653978
2164098712464783467218987639450897764398217346946
4964392764309872634287469864987597152397123976490
8714698764987243698127346987461435895321456865437

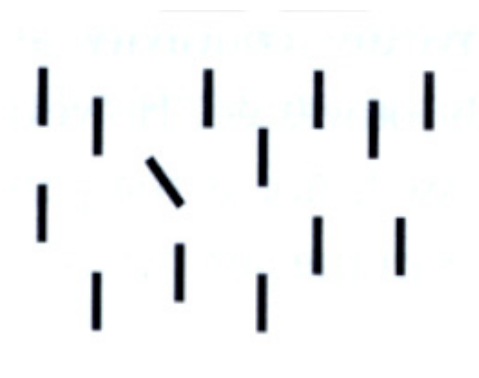
Typical Results



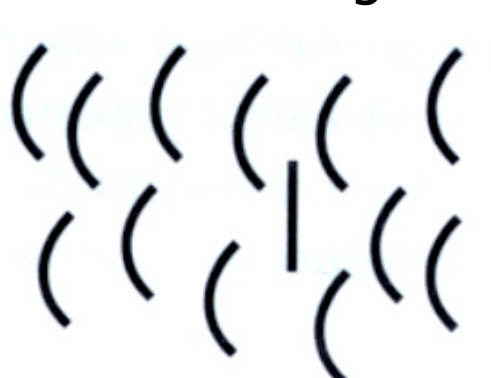
Pre-attentive processing

68

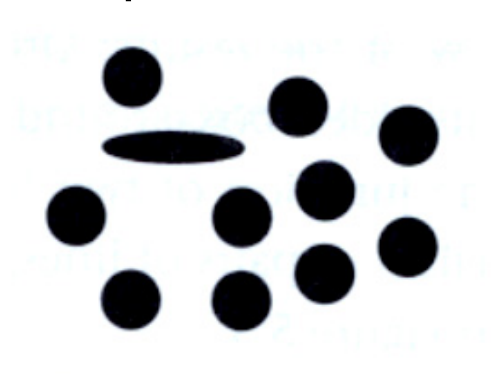
orientation



curved/straight



shape



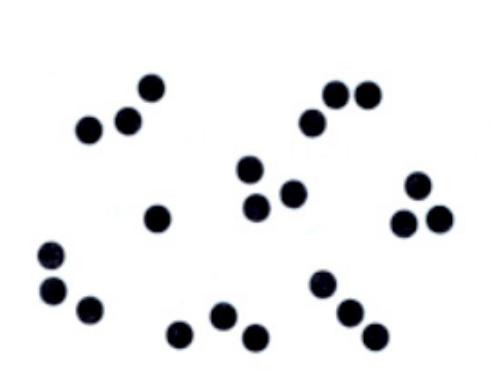
shape



size



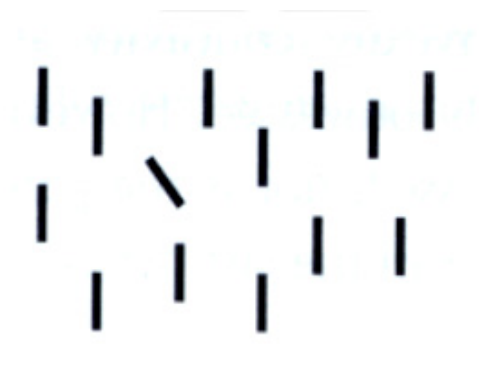
number



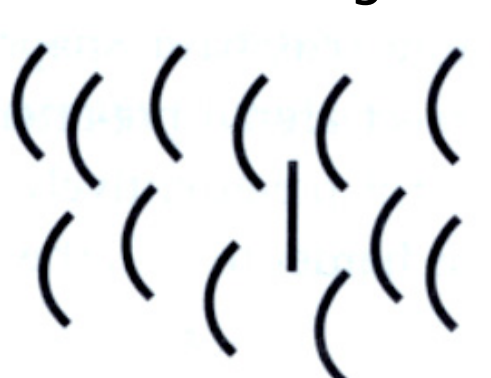
Pre-attentive processing

69

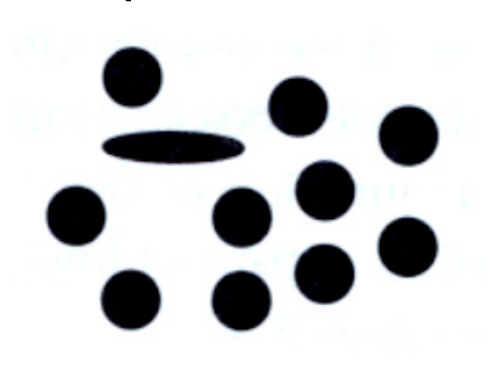
orientation



curved/straight



shape



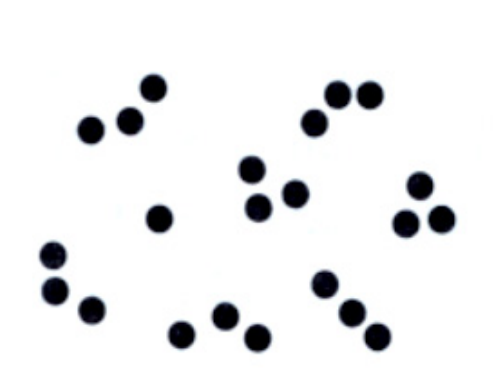
shape



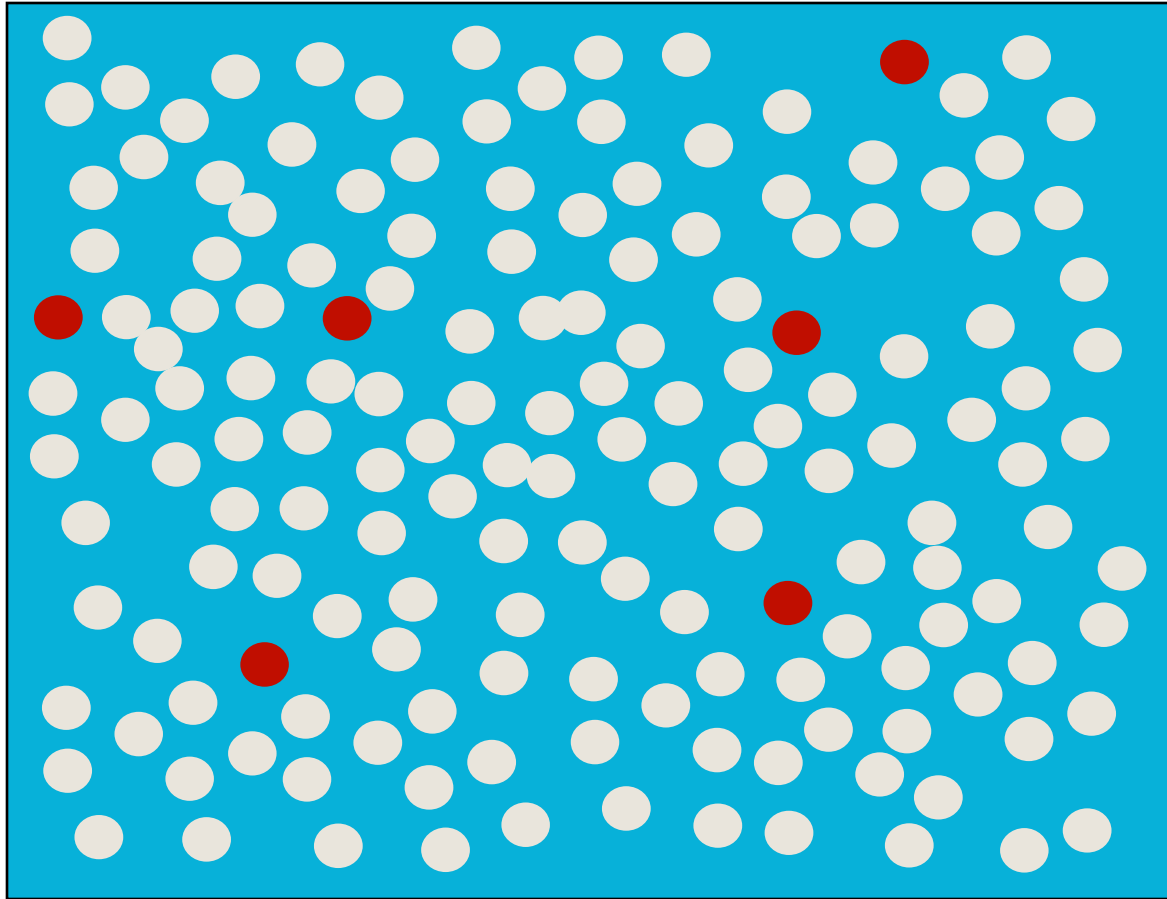
size



number



Colour



Shape

