

Foundations 1

Basic Concepts

DataVis 2020

<http://datavis2020.github.io>

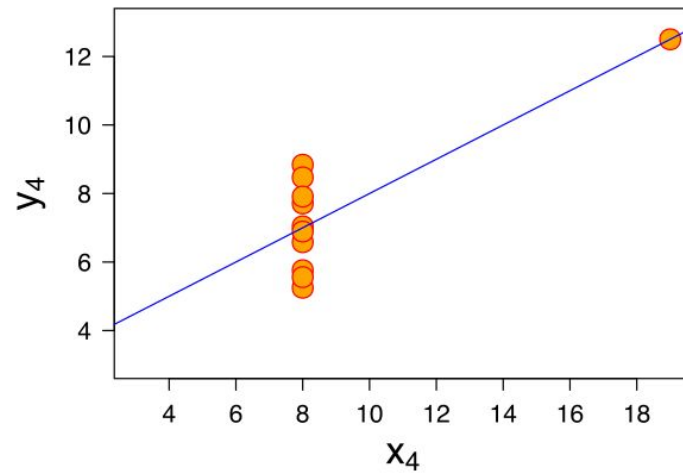
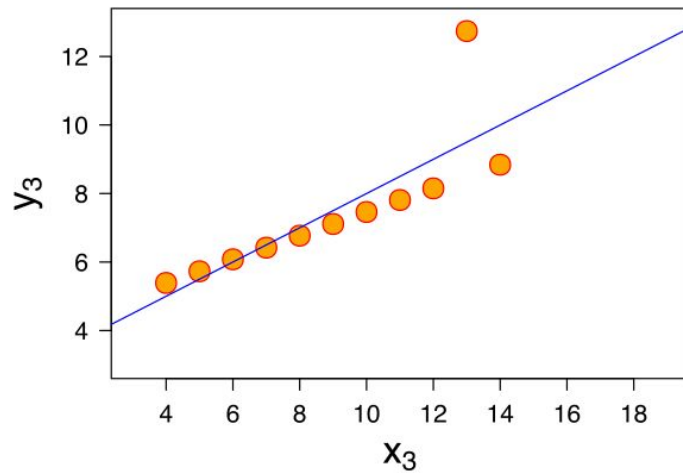
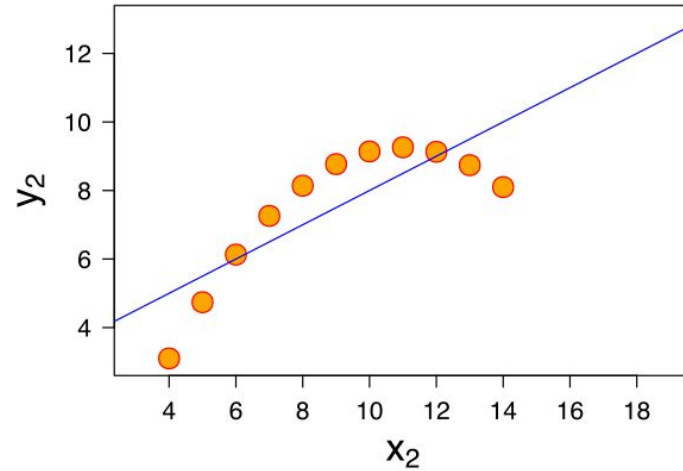
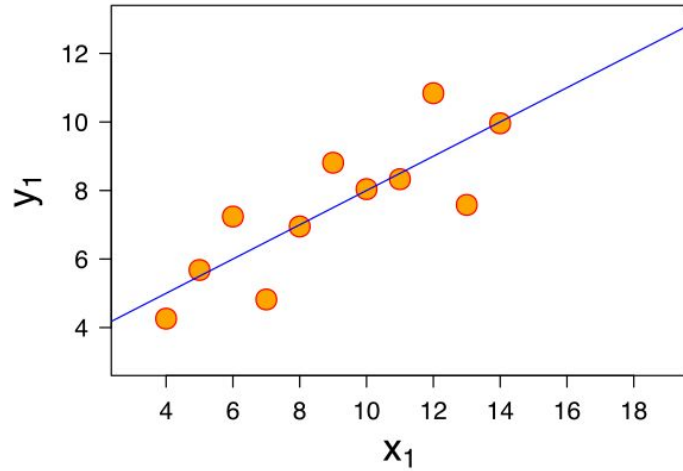
Dr. Benjamin Bach



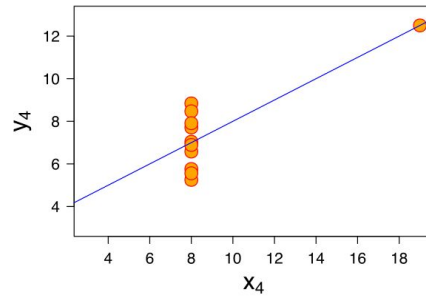
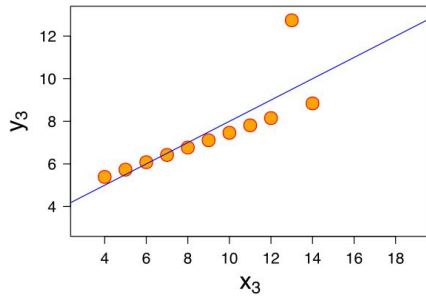
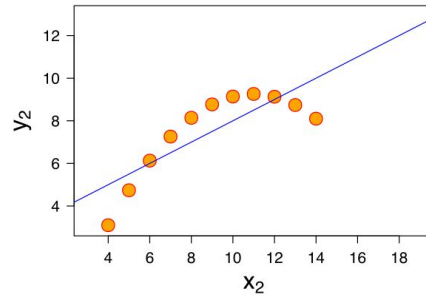
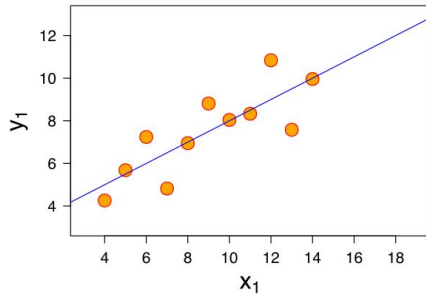
THE UNIVERSITY
of EDINBURGH

What is visualization?

Anscombe's Quartet



Anscombe's Quartet



Property	Value
Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression	0.67

(Possible) Definitions

"The use of **computer-supported, interactive, visual representations** of **abstract data** to amplify **cognition**"

Stuart Card

..an **accessible way** to **see** and **understand trends, outliers, and patterns** in data."

Tableau

... to help **people** carry out **tasks** more **effectively**

Tamara Munzner

Defining Concepts

- computer supported
- Interactive visual representations
- For Abstract data
- Helping people
- to see and understand
- trends, outliers, and patterns in data,
- and carry out tasks
- more effectively
- Through amplifying cognition

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box		7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69		5 4-Not Specified	Small Pack	0.44	6/6/05
69		5 4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06

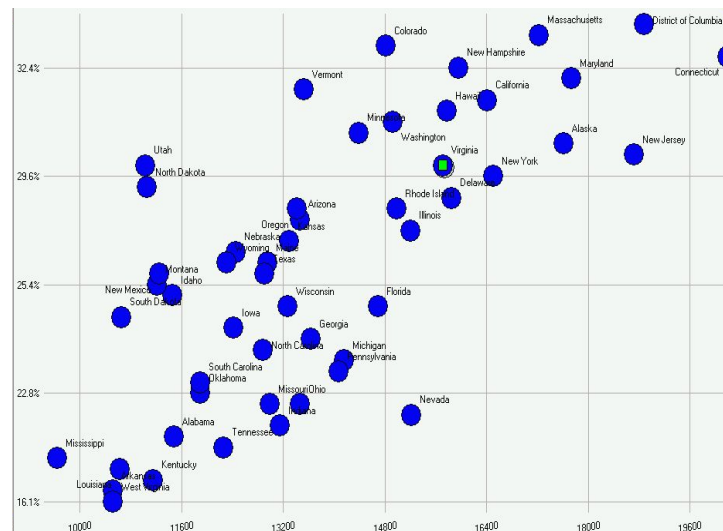
attribute

item

cell



Data → Visualization → Information → Action



Encoding: designer

1

Data

What is my data?

Which data type?

Ordinal / numerical / categorical?

2

Visual Mapping

What is my visual representation?

Which visual variables am I using?

How am I encoding my data?

3

Rendering

What is my medium?

monoscopic/stereoscopic?

Tangibility?

Print / digital?

View

Decoding: user

6

Comprehending

What does it mean for me?

What shall I do now?

Is this all true?

What do I learn?

5

Interpreting

What does it mean?

What does color mean?

What does 'up' mean?

What do these patterns show?

4

Perceiving

What does it show?

Where is big, medium, small?

How do things compare?

What relationships exist?

1. Data

Value Attribute

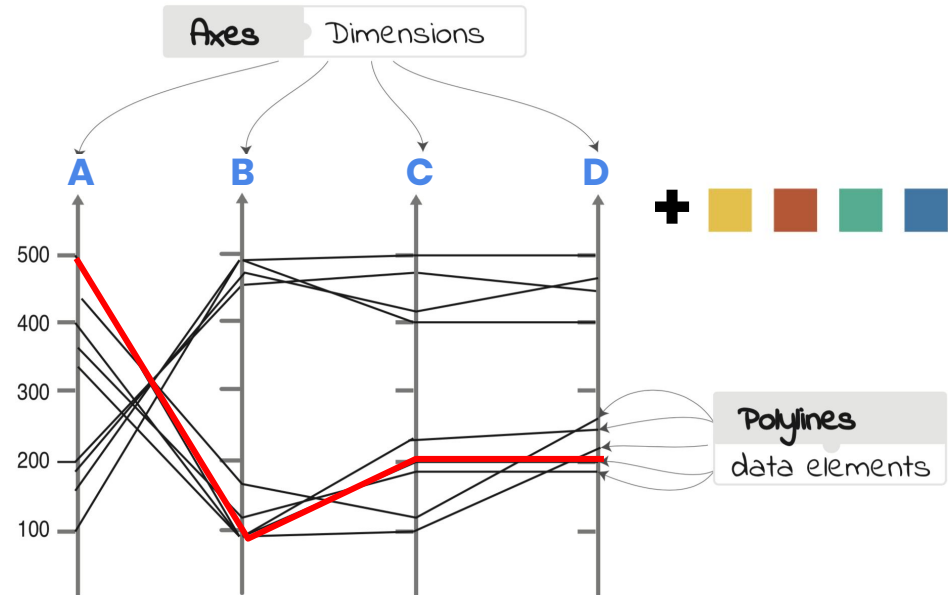
1		mpg	cyl	disp	hp	drat	wt	qsec
2	Mazda RX4	21	6	160	110	3.9	2.62	16.46
3	Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02
4	Datsun 710	22.8	4	108	93	3.85	2.32	18.61
5	Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44
6	Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02
7	Valiant	18.1	6	225	105	2.76	3.46	20.22
8	Duster 360	14.3	8	360	245	3.21	3.57	15.84
9	Merc 240D	24.4	4	146.7	62	3.69	3.19	20
10	Merc 230	22.8	4	140.8	95	3.92	3.15	22.9
11	Merc 280	19.2	6	167.6	123	3.92	3.44	18.3
12	Merc 280C	17.8	6	167.6	123	3.92	3.44	18.9
13	Merc 450SE	16.4	8	275.8	180	3.07	4.07	17.4
14	Merc 450SL	17.3	8	275.8	180	3.07	3.73	17.6
15	Merc 450SLC	15.2	8	275.8	180	3.07	3.78	18
16	Cadillac Fleetwood	10.4	8	472	205	2.93	5.25	17.98

Item

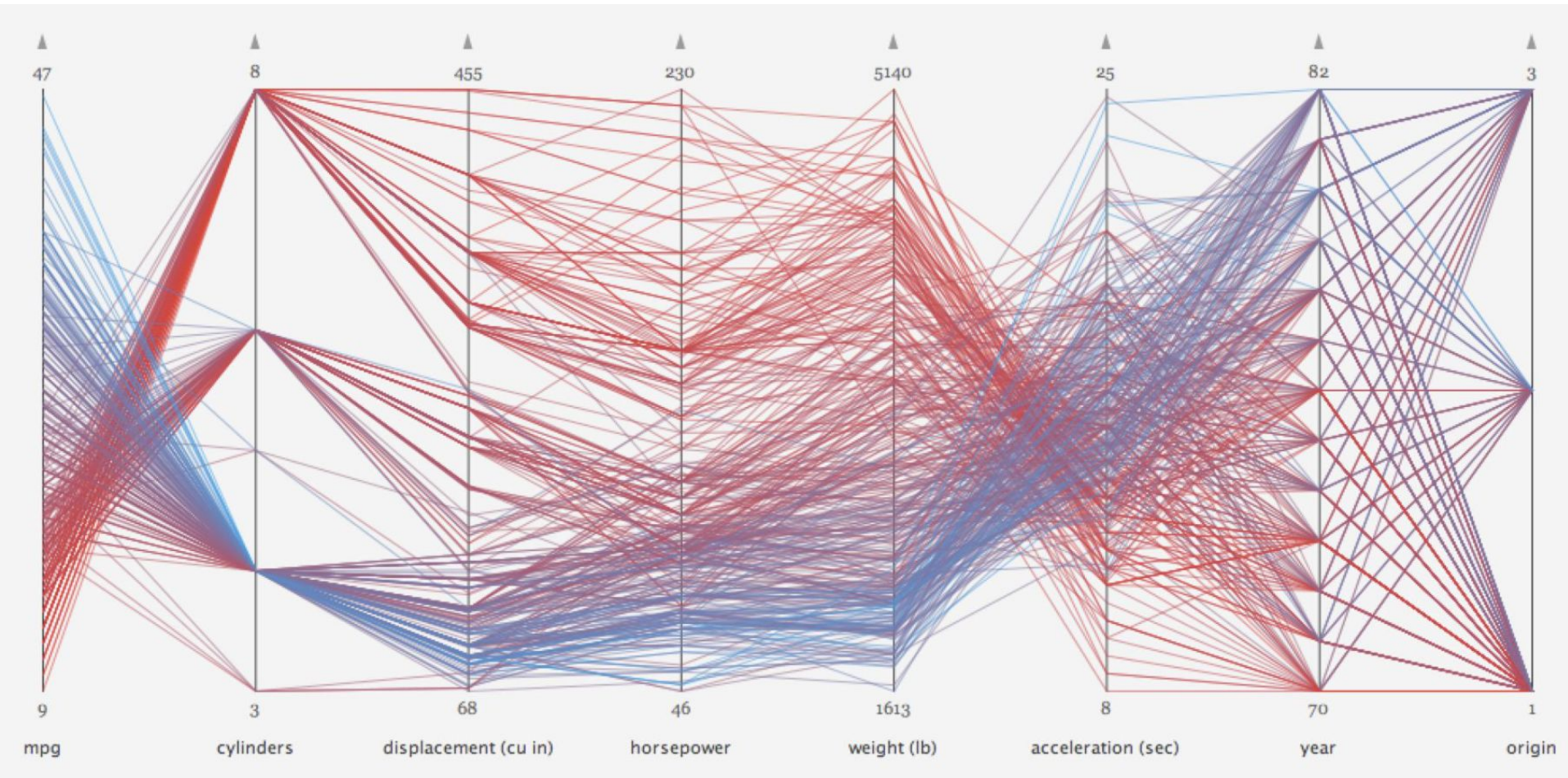
2. Visual Mapping

1		mpg	cyl	disp	hp	drat	wt	qsec
2	Mazda RX4	21	6	160	110	3.9	2.62	16.46
3	Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02
4	Datsun 710	22.8	4	108	93	3.85	2.32	18.61
5	Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44
6	Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02
7	Valiant	18.1	6	225	105	2.76	3.46	20.22
8	Duster 360	14.3	8	360	245	3.21	3.57	15.84
9	Merc 240D	24.4	4	146.7	62	3.69	3.19	20
10	Merc 230	22.8	4	140.8	95	3.92	3.15	22.9
11	Merc 280	19.2	6	167.6	123	3.92	3.44	18.3
12	Merc 280C	17.8	6	167.6	123	3.92	3.44	18.9
13	Merc 450SE	16.4	8	275.8	180	3.07	4.07	17.4
14	Merc 450SL	17.3	8	275.8	180	3.07	3.73	17.6
15	Merc 450SLC	15.2	8	275.8	180	3.07	3.78	18
16	Cadillac Fleetwood	10.4	8	472	205	2.93	5.25	17.98

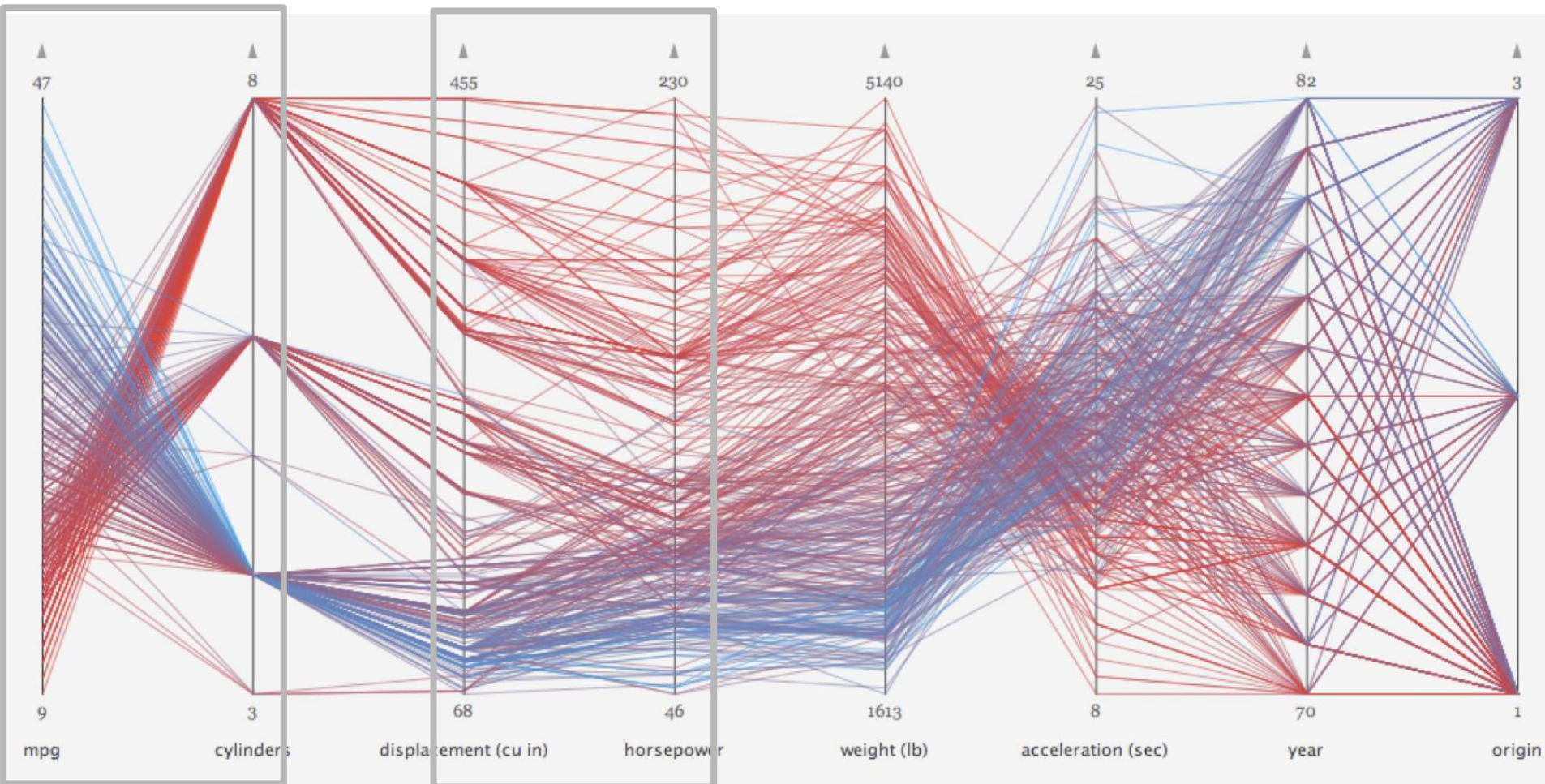
Data



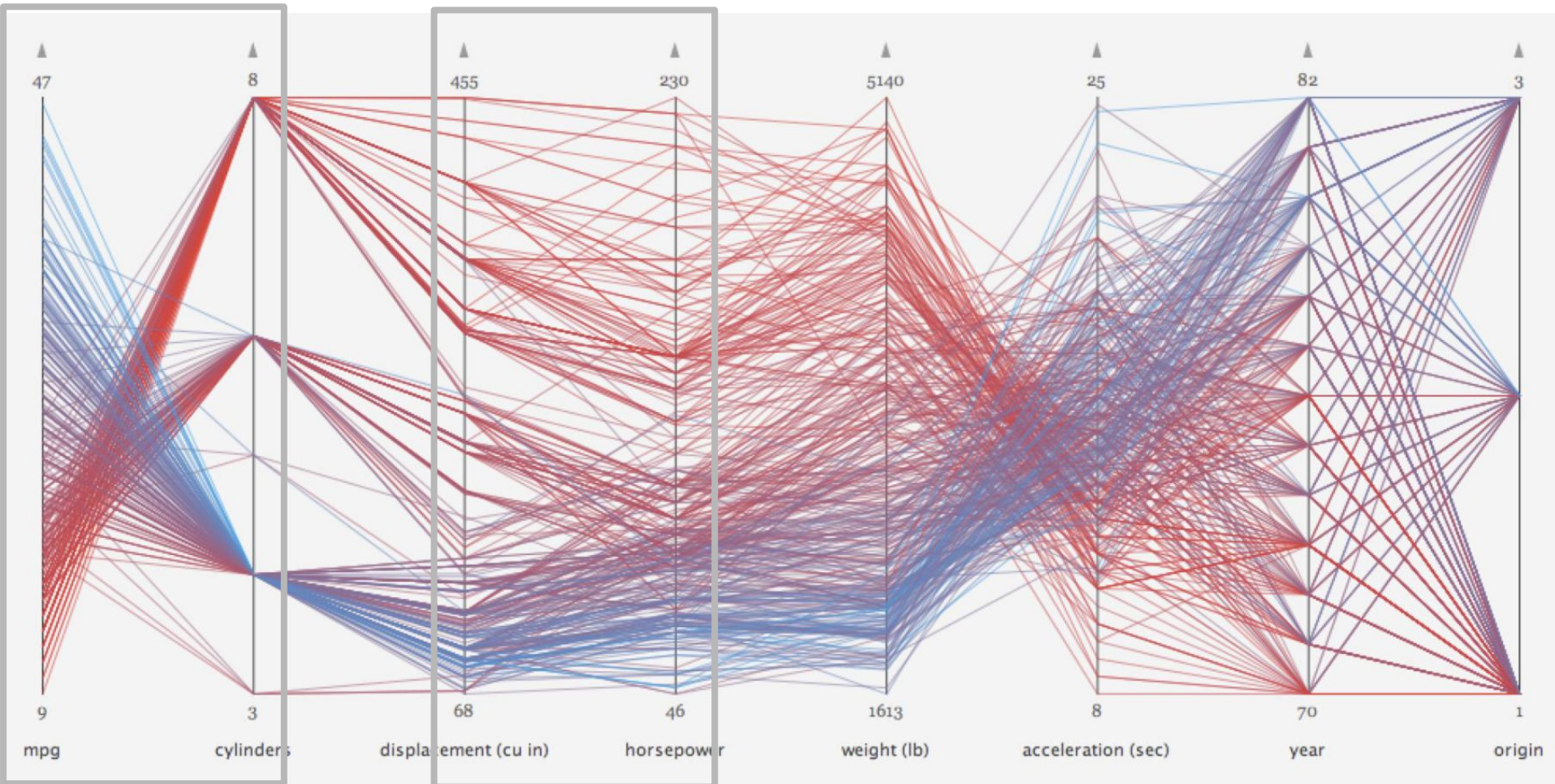
3. Rendering



4. Perceiving



5. Interpretation



**Negative
Correlation**



**Positive
Correlation**

Encoding: designer

1

Data

What is my data?

Which data type?

Ordinal / numerical / categorical?

2

Visual Mapping

What is my visual representation?

Which visual variables am I using?

How am I encoding my data?

3

Rendering

What is my medium?

monoscopic/stereoscopic?

Tangiblity?

Print / digital?

View

Decoding: user

6

Comprehending

What does it mean for me?

What shall I do now?

Is this all true?

What do I learn?

5

Interpreting

What does it mean?

What does color mean?

Wha does 'up 'mean?

What do these patterns show?

4

Perceiving

What does it show?

Where is big, medium, small?

How do things compare?

What relationships exist?

Find every '5'

9176867991960386255930486551443
9353652502752141394912668766013
5095444663473545993604636448078
6425620171155906710121069218701
0584448205561385902503573068845
1172102391879794238892669764599
6987227951015948926017759166604
8436204036527029291707742717051
3280506428553158863136868380421
4055135906770329783671748392874

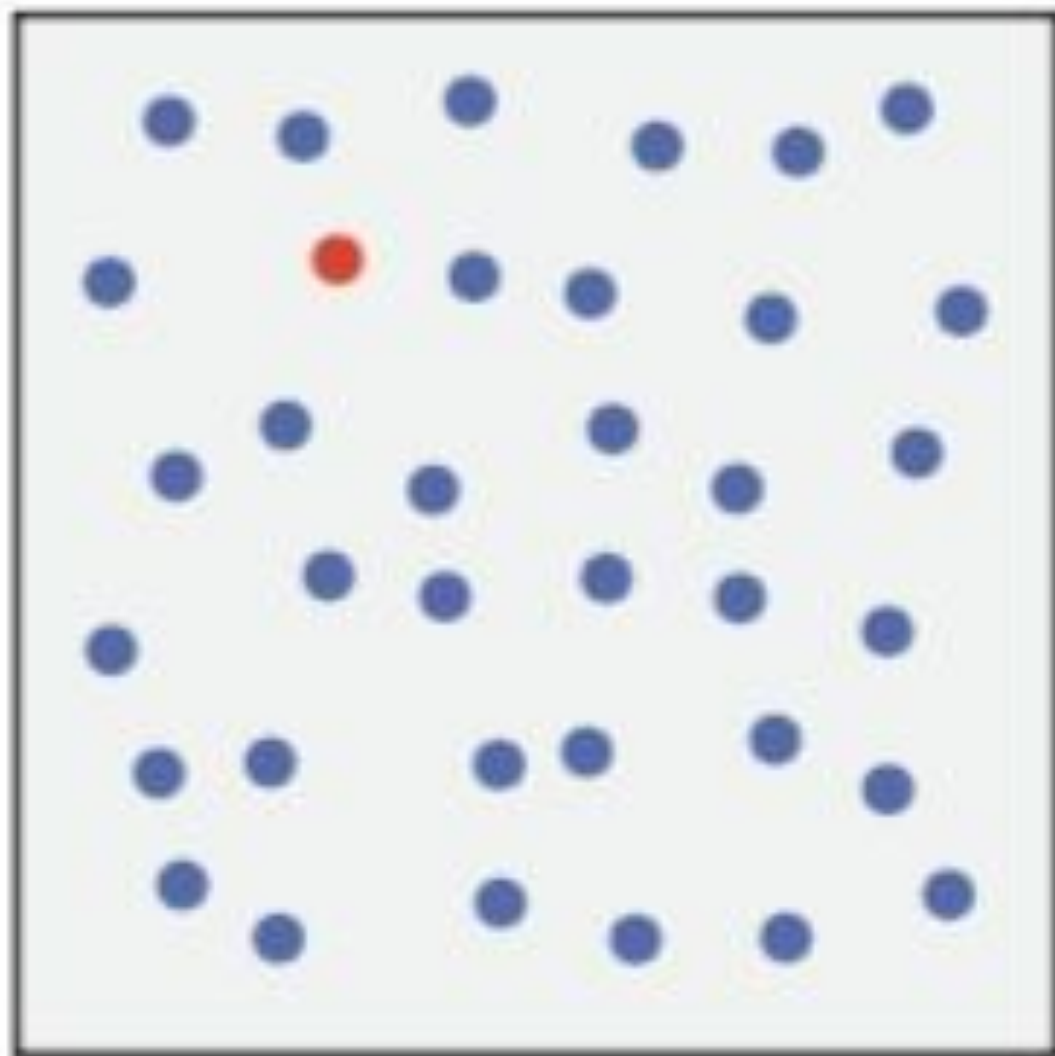
Pre-attentiveness:

- *"is the subconscious accumulation of information from the environment"*
- realizing something before you think

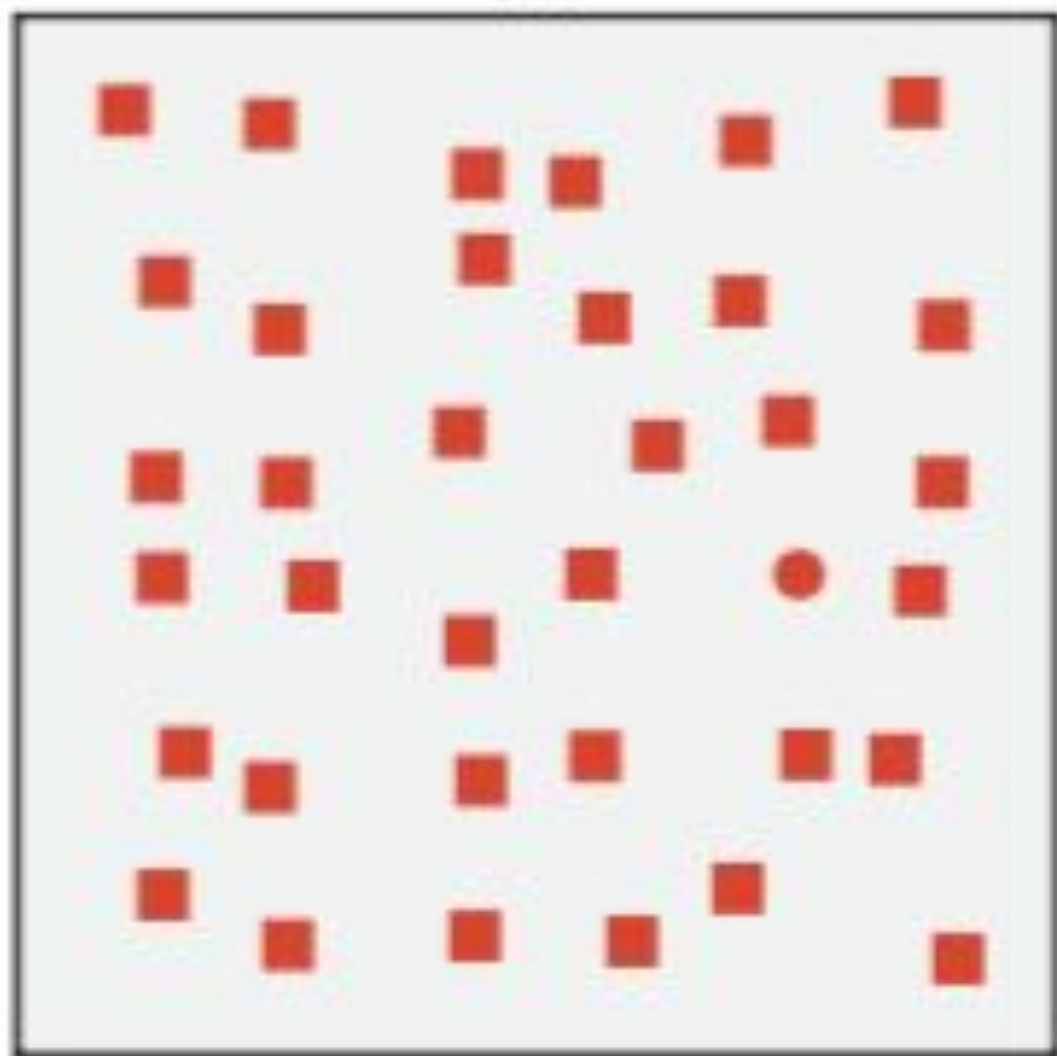
Parallelity:

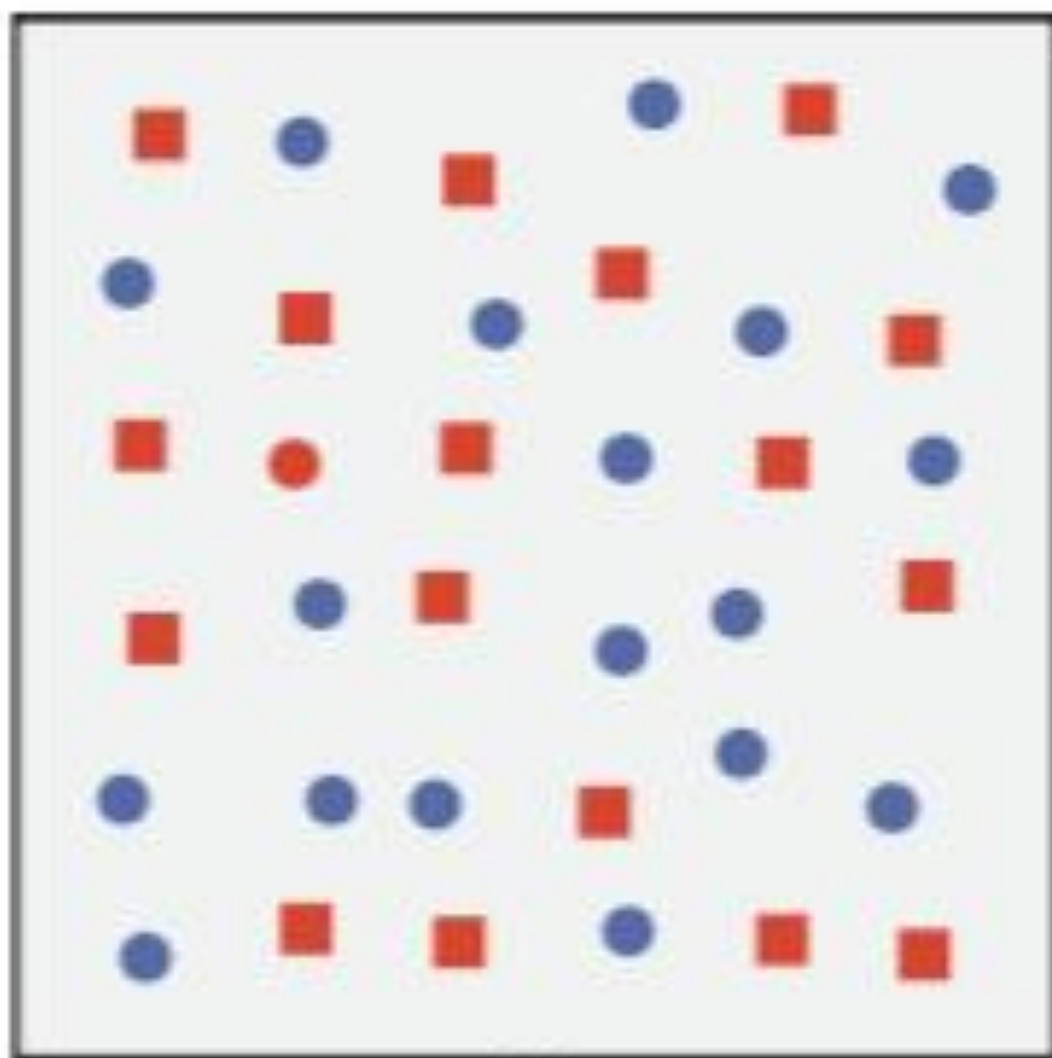
- Perceiving stimuli in parallel

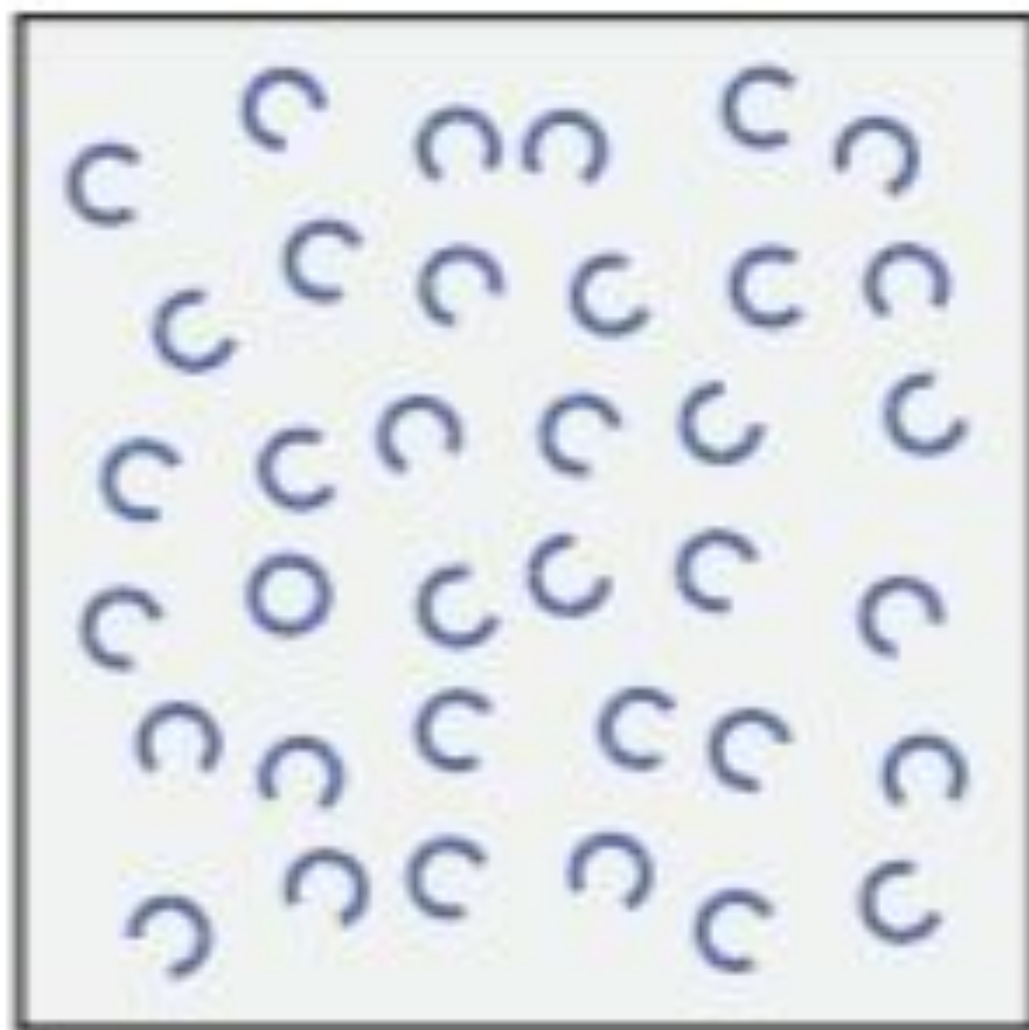




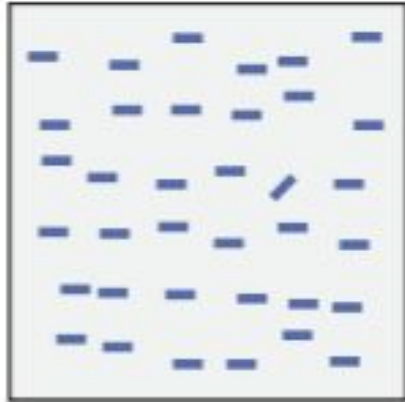
8-7



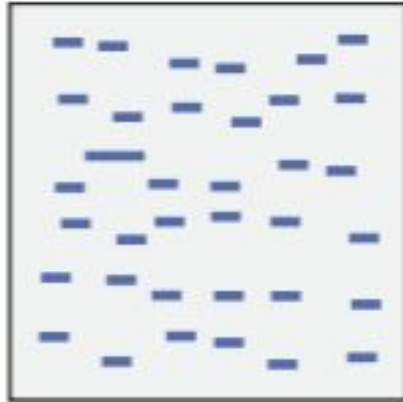




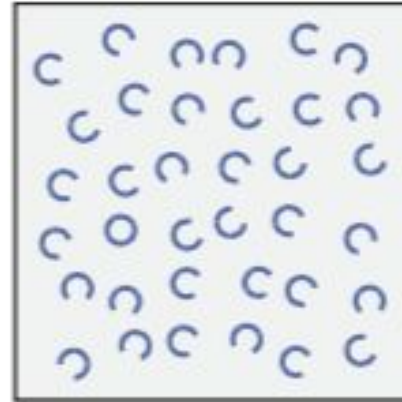
Preattentiveness



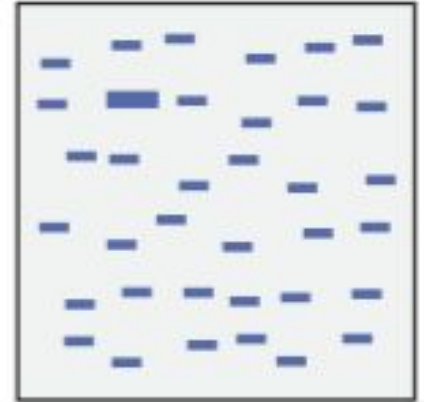
orientation
[16], [17], [18], [19]



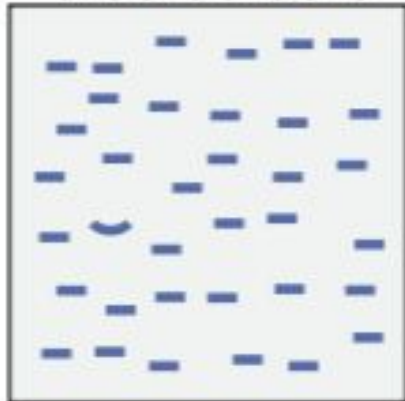
length
[20], [21]



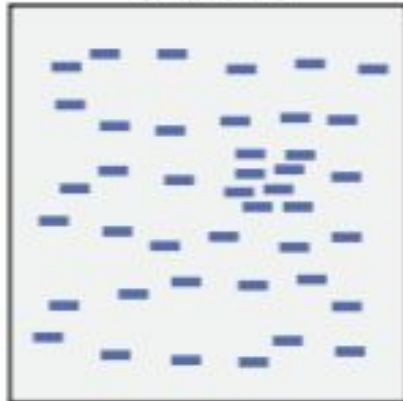
closure
[16]



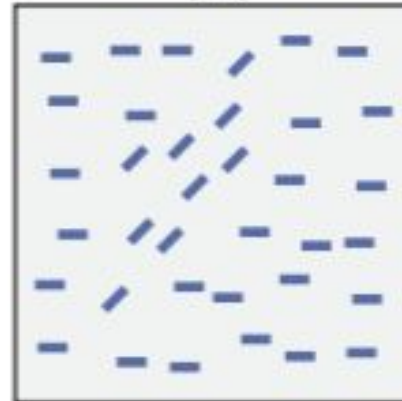
size
[22], [23]



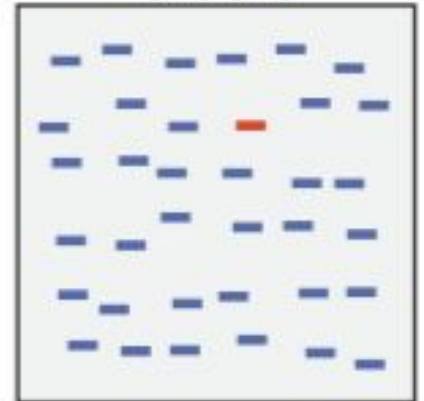
curvature
[21]



density
[23]



number
[20], [24], [25]



hue
[23], [26], [27], [28], [29]

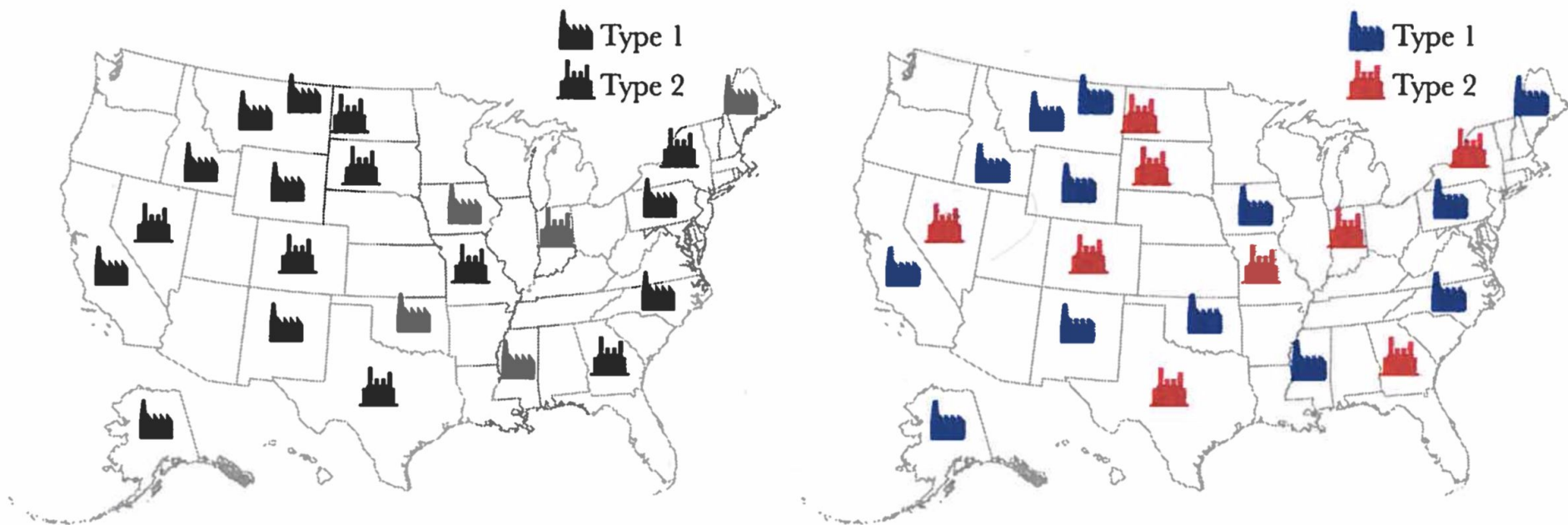
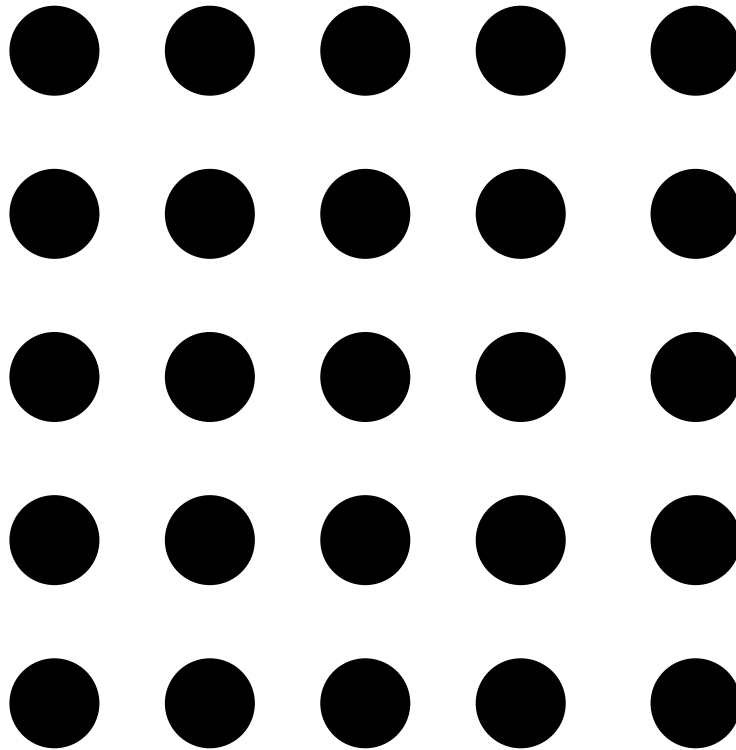
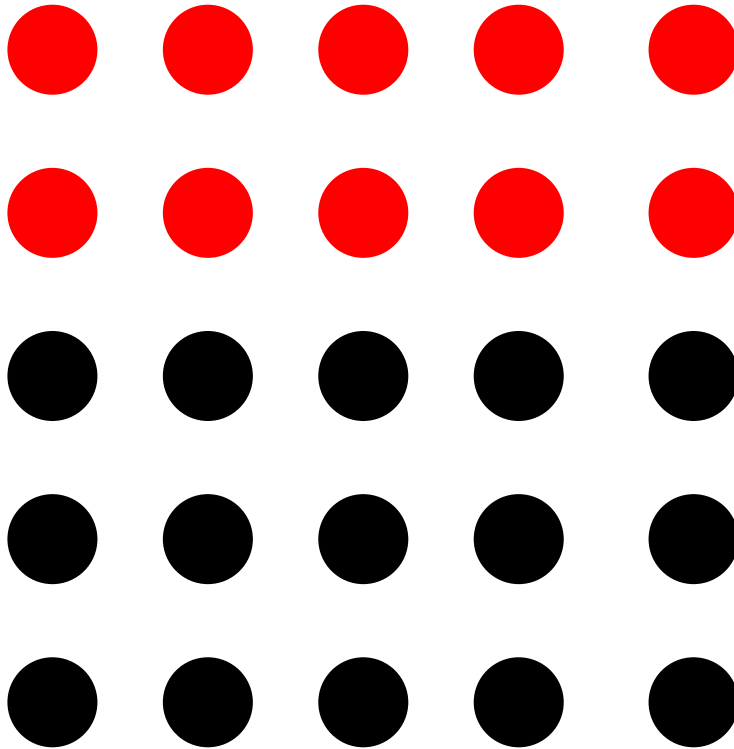


Figure 6.3 On which of these maps is it easier to identify the number of factories of each kind?

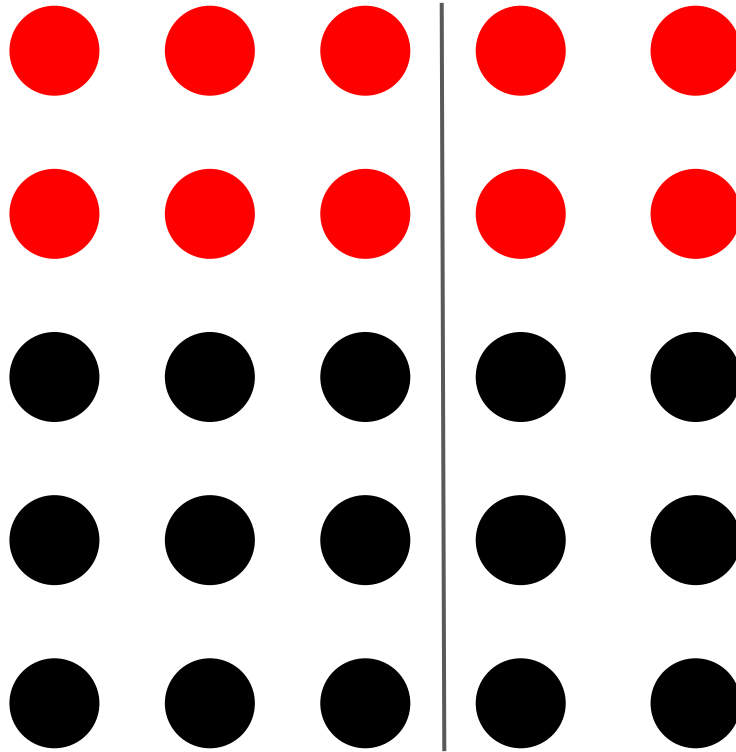
Gestalt Laws



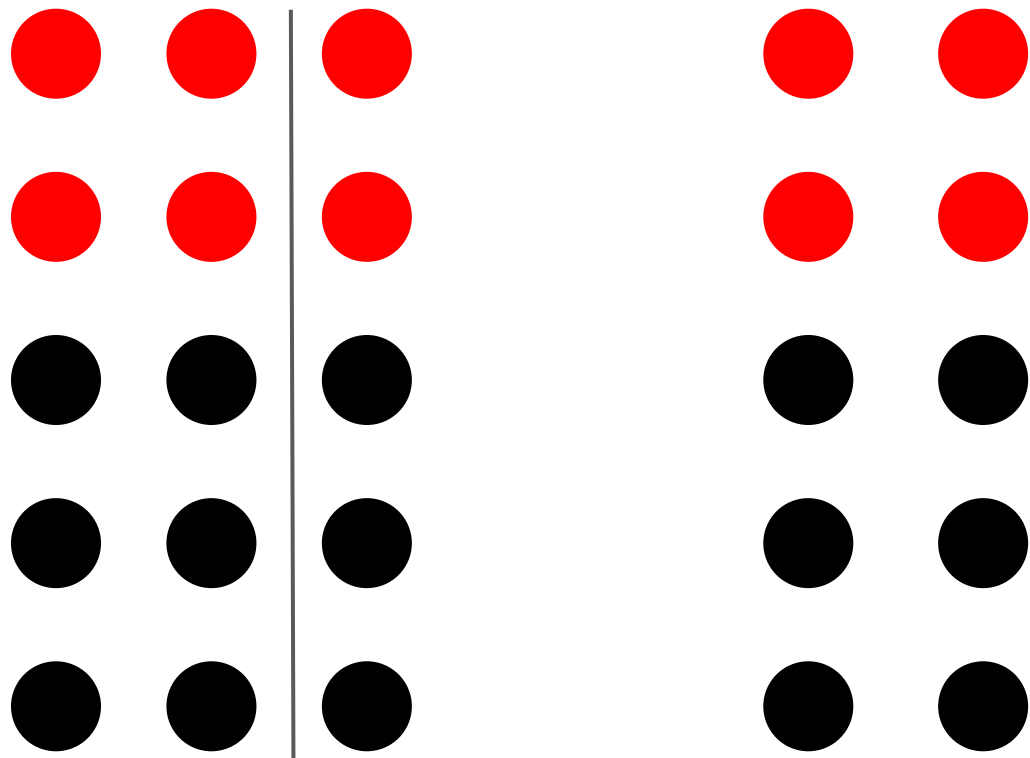
Law of Similarity



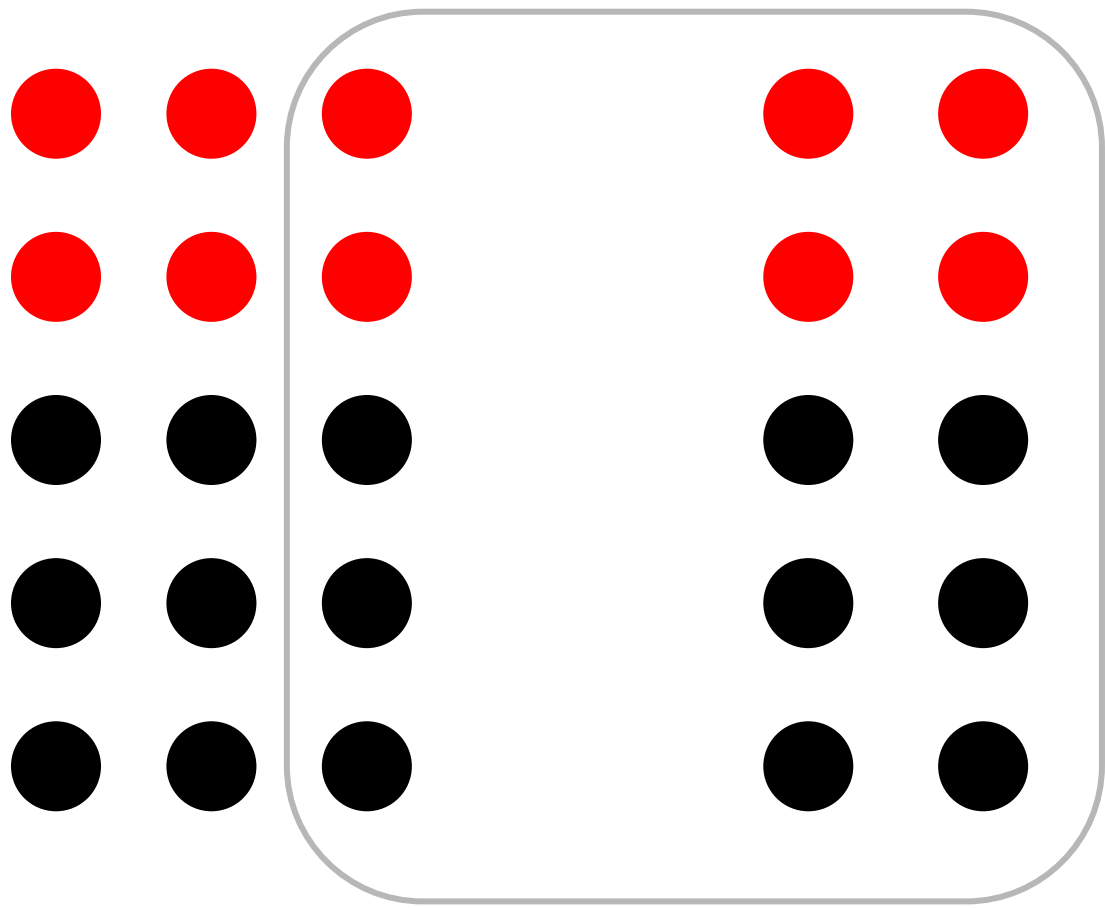
Law of Closure



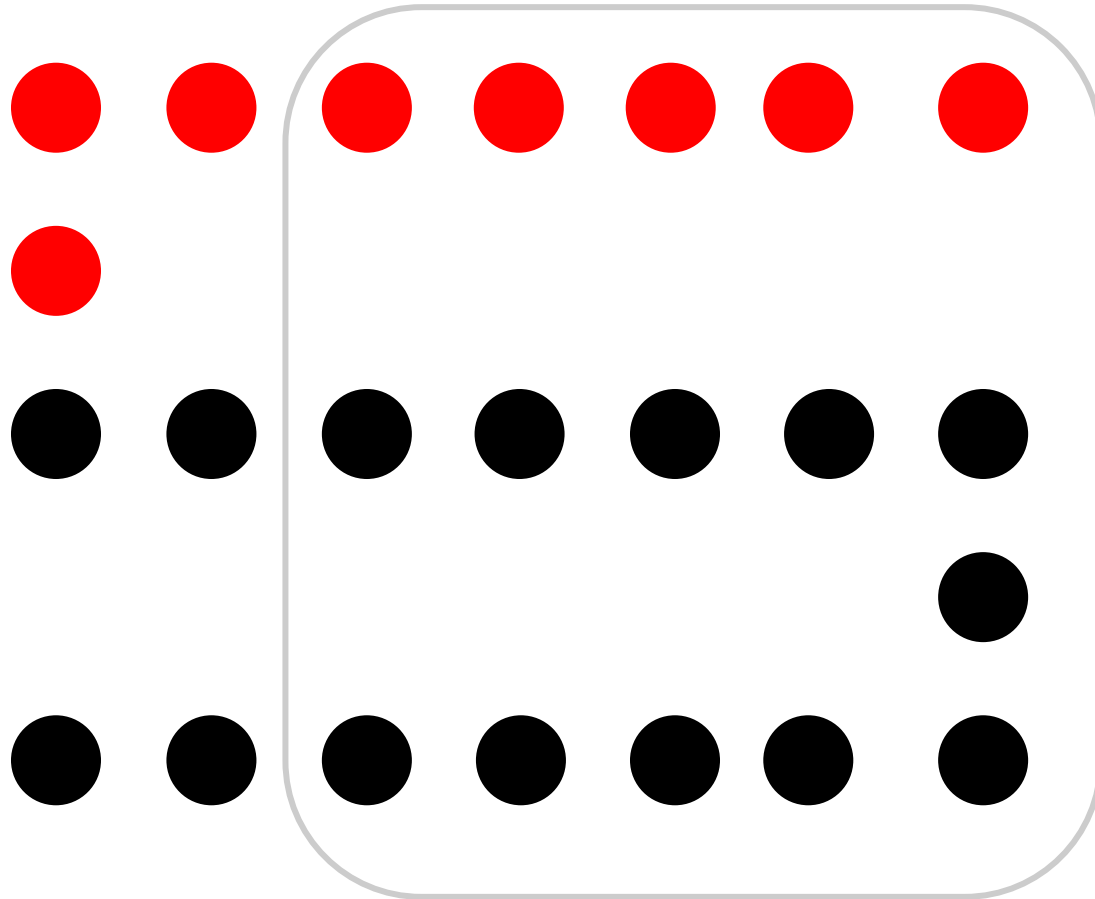
Law of Proximity



Law of Enclosure

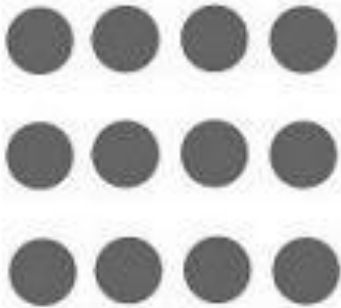


Law of Continuity

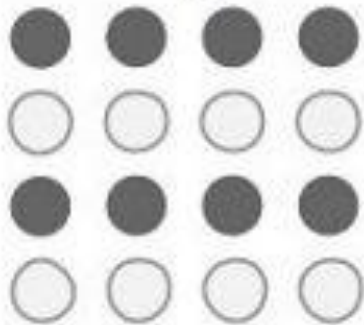


Gestalt Laws

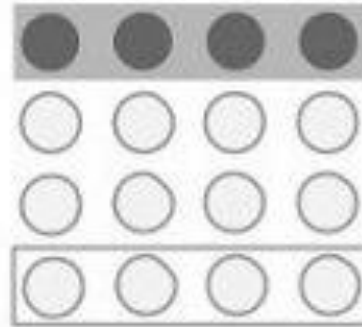
Proximity



Similarity



Enclosure



Symmetry



Closure



Continuity



Connection

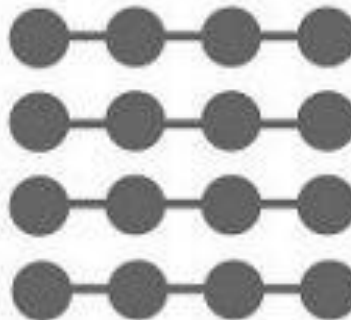


Figure & ground



Gestalt Laws



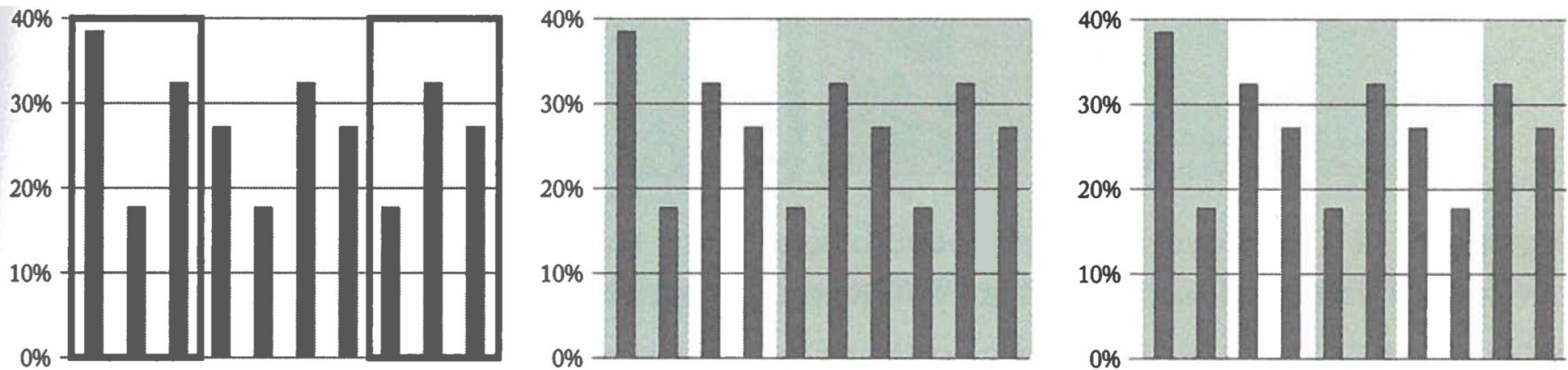


Figure 6.10 Boxing bars helps readers identify groups.

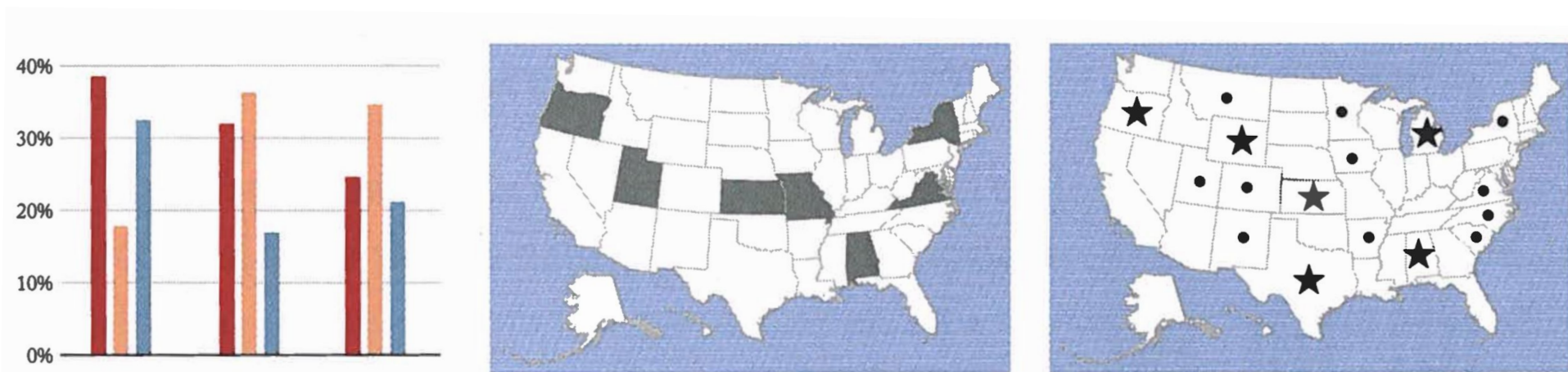
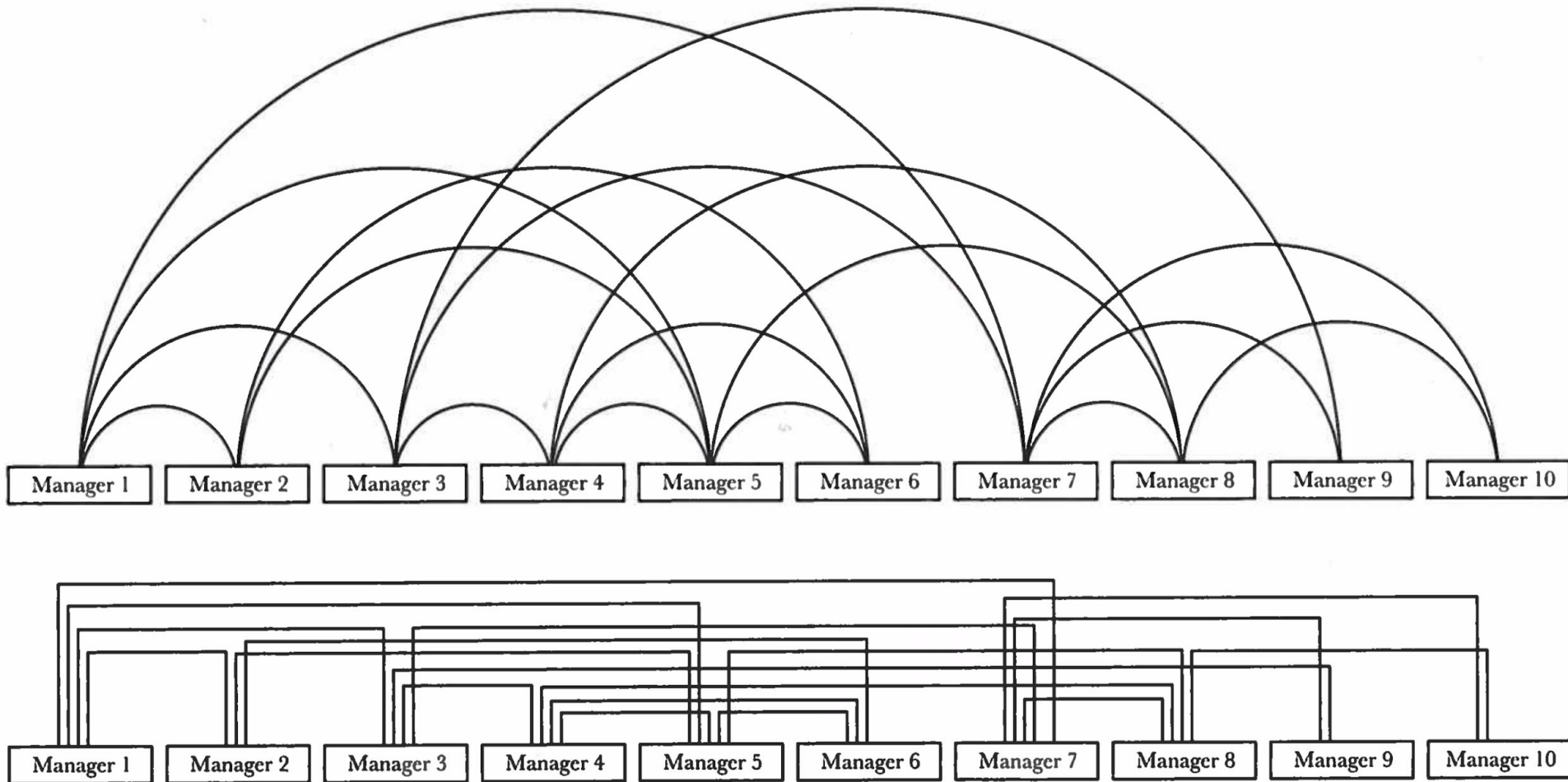


Figure 6.7 Objects that look alike will be identified as parts of a group.

Continuity for line following

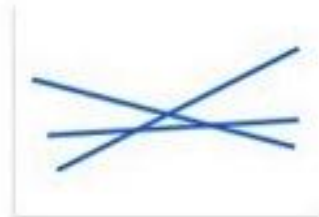


Visual Patterns

High, low and in between



Going up, going down and remaining flat



Steep and gradual



Visual Patterns

Steady and fluctuating



Random and repeating



Straight and curved



Encoding: designer

1

Data

What is my data?

Which data type?

Ordinal / numerical / categorical?

2

Visual Mapping

What is my visual representation?

Which visual variables am I using?

How am I encoding my data?

3

Rendering

What is my medium?

monoscopic/stereoscopic?

Tangibility?

Print / digital?

View

Decoding: user

6

Comprehending

What does it mean for me?

What shall I do now?

Is this all true?

What do I learn?

5

Interpreting

What does it mean?

What does color mean?

What does 'up' mean?

What do these patterns show?

4

Perceiving

What does it show?

Where is big, medium, small?

How do things compare?

What relationships exist?

Data Types

Categorical
(*Qualitative*)



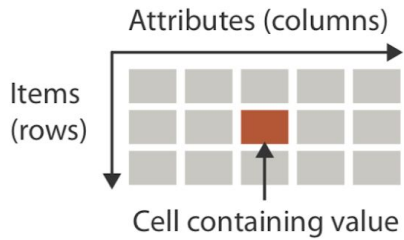
Ordered

Numerical
(*Quantitative*)

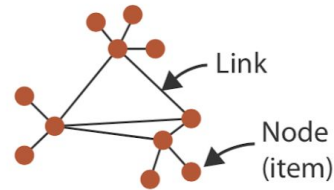


Dataset Types

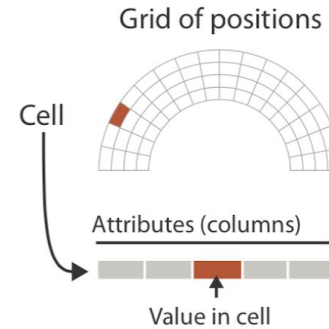
→ Tables



→ Networks



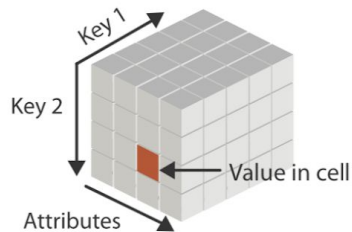
→ Fields (Continuous)



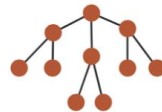
→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



Encoding: designer

1

Data

What is my data?

Which data type?

Ordinal / numerical / categorical?

2

Visual Mapping

What is my visual representation?

Which visual variables am I using?

How am I encoding my data?

3

Rendering

What is my medium?

monoscopic/stereoscopic?

Tangiblity?

Print / digital?

View

Decoding: user

6

Comprehending

What does it mean for me?

What shall I do now?

Is this all true?

What do I learn?

5

Interpreting

What does it mean?

What does color mean?

Wha does 'up 'mean?

What do these patterns show?

4

Perceiving

What does it show?

Where is big, medium, small?

How do things compare?

What relationships exist?

Visual Marks



Points



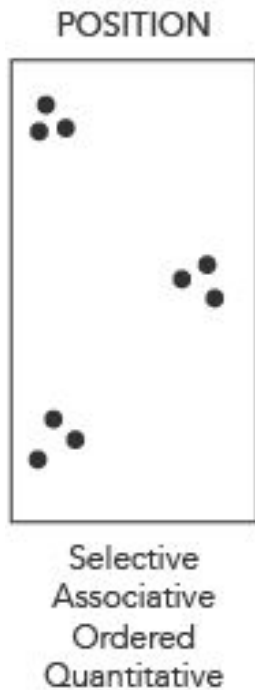
Lines



Areas

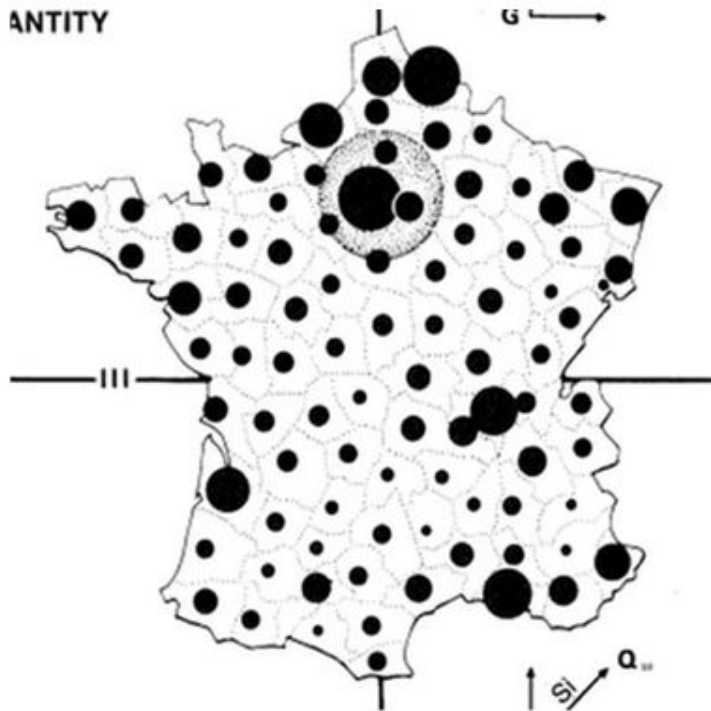
Visual Variables

Bertin's Visual Variables

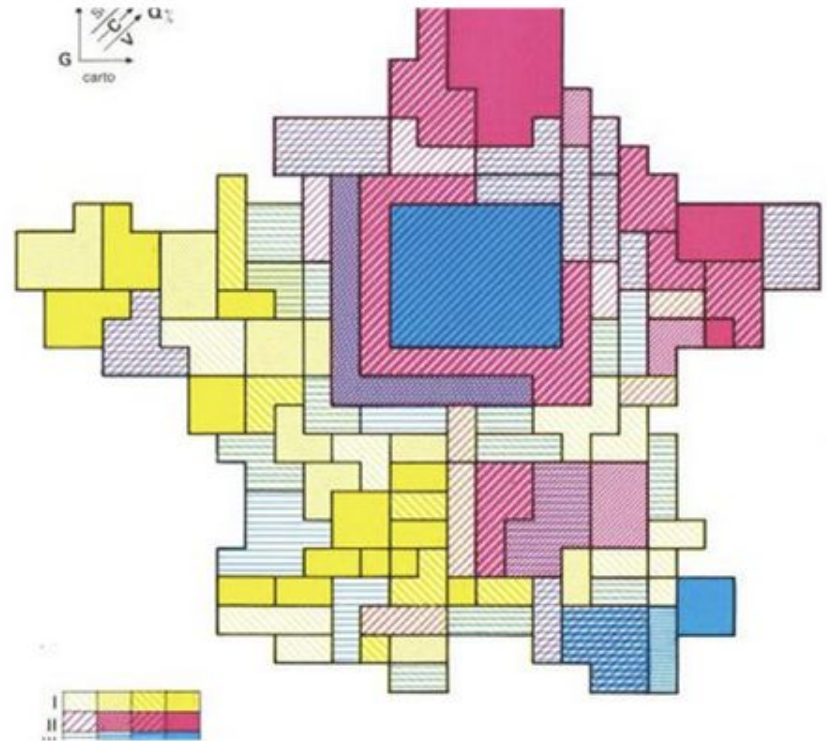


Visual Variables

ANTITY



- Points
- Size



- Areas
- Size
- Hue
- Texture

Characteristics of Visual Variables

Selective: *Does it make a mark distinguishable?*

Associative: *Can I group marks?*

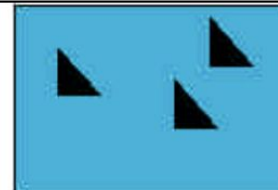
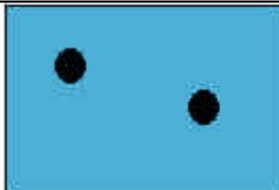
Quantitative: *Can I do maths with it?*

Order: *Does it imply an order on the marks?*

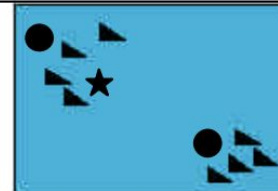
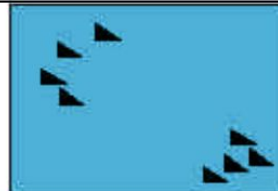
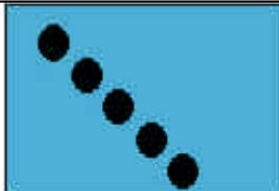
Length: *How many values does the variable has?*

Visual Variable: Position

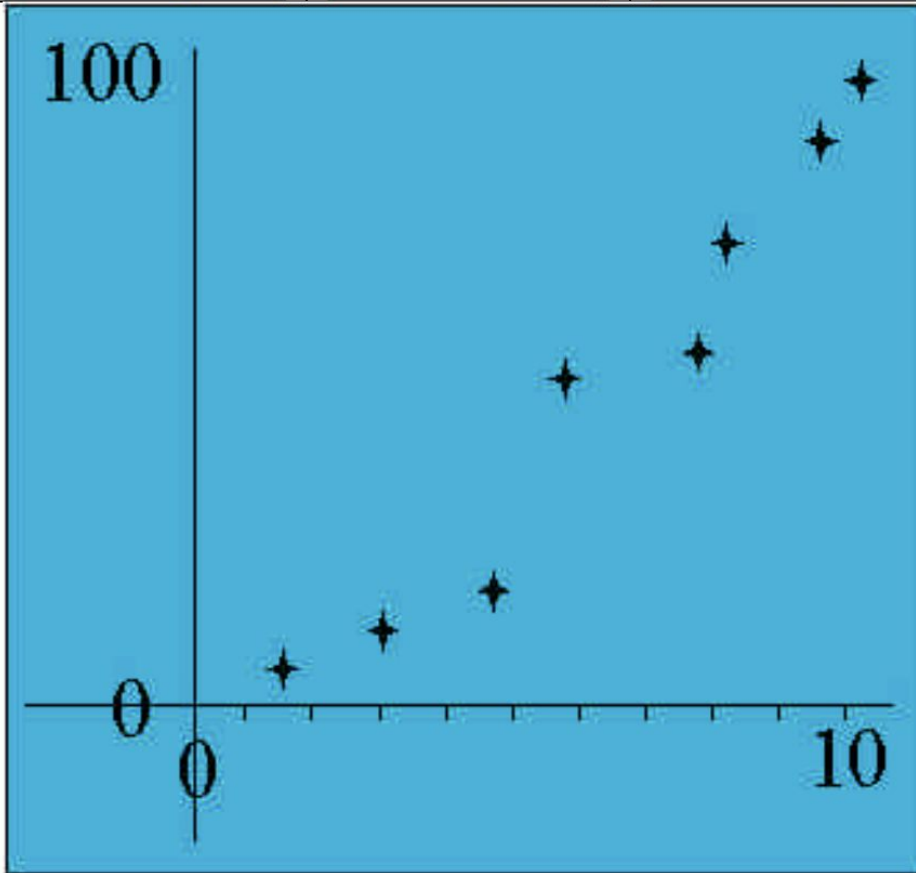
selective



associative




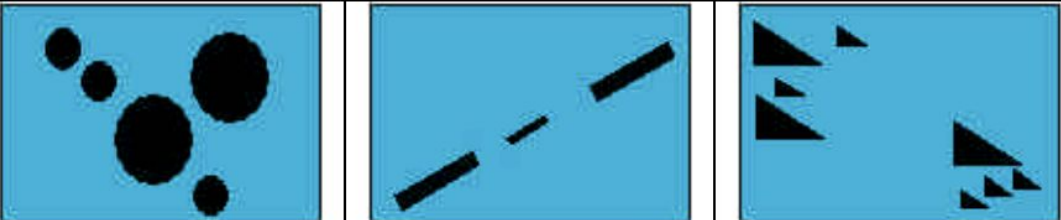
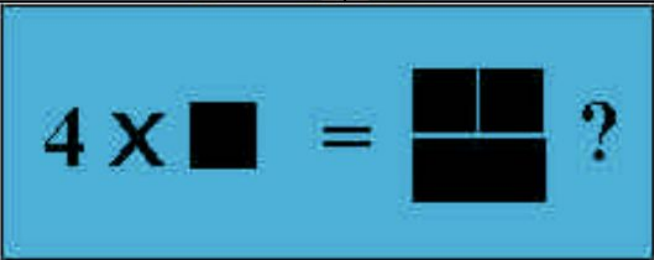
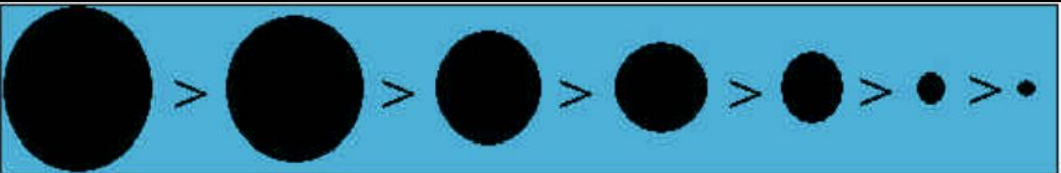
quantitative



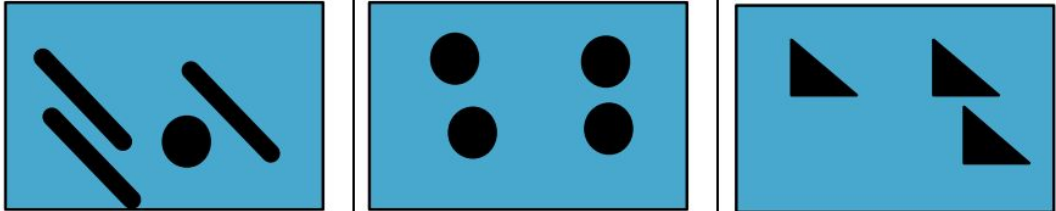
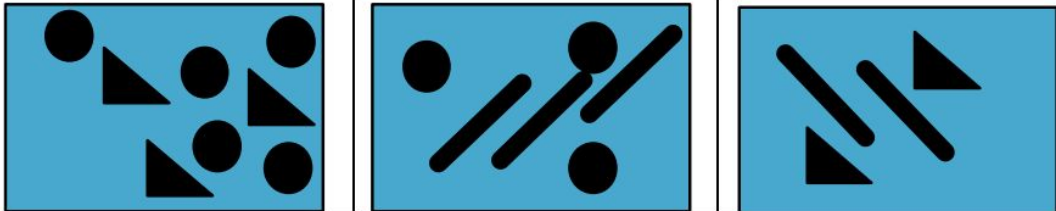


order

length

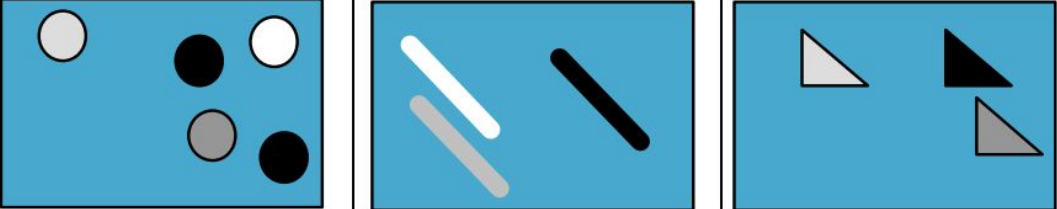

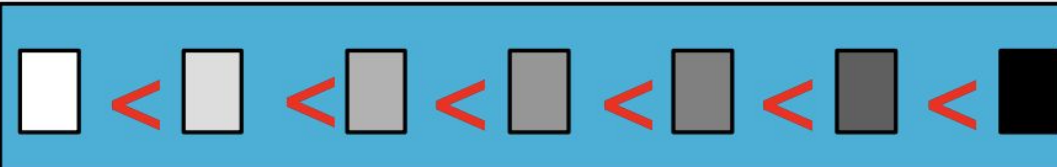
Visual Variable: Size

	selective	
	associative	
	quantitative	
	order	
	Length	<ul style="list-style-type: none"> • theoretically infinite but practically limited • association and selection ~ 5 and distinction ~ 20

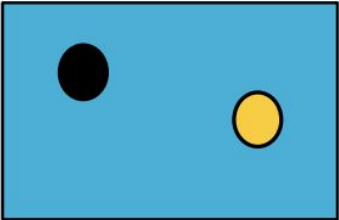
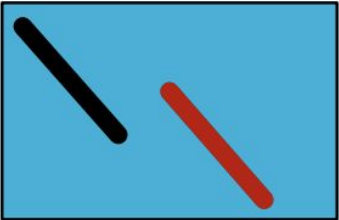
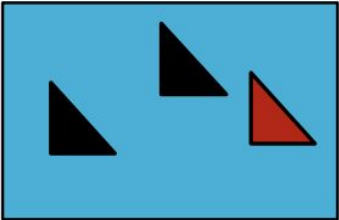
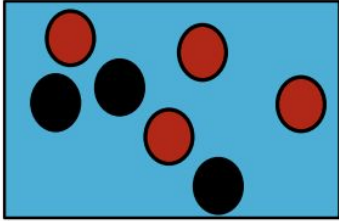
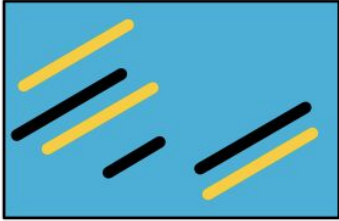
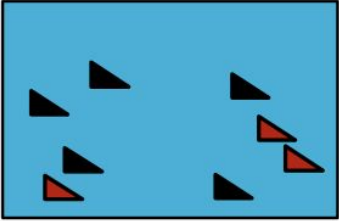
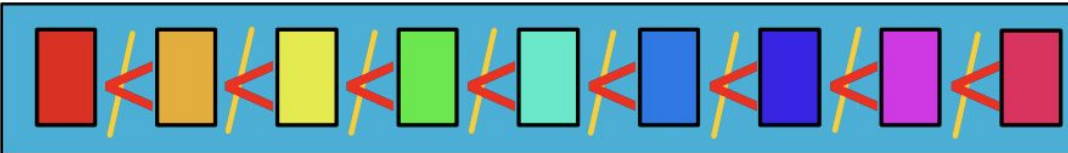

Visual Variable: Shape

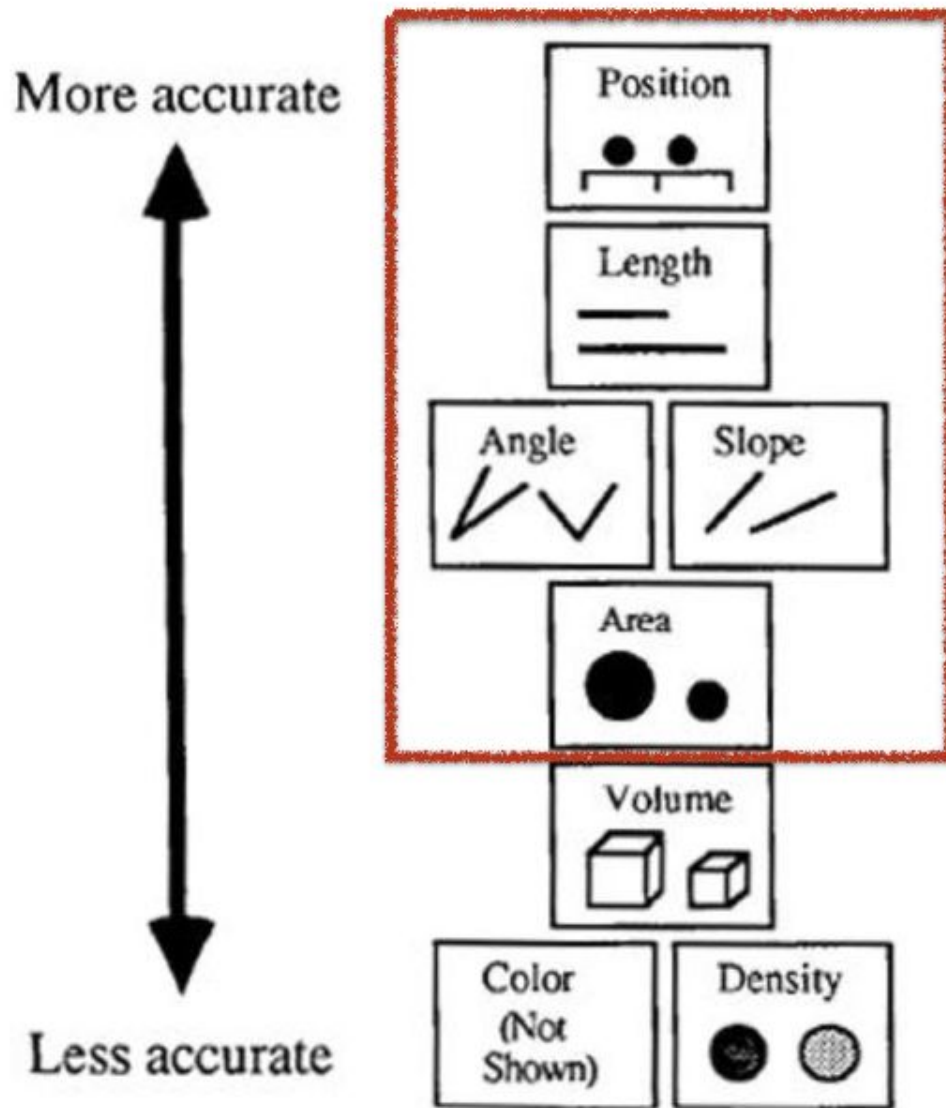
	selective	
	associative	
	quantitative	
	order	
	length	 <p>theoretically infinite</p>

Visual Variable: Value

✓	selective	
✓	associative	
≠	quantitative	
✓	order	
✓	length	<ul style="list-style-type: none"> theoretically infinite but practically limited association and selection ~ < 7 and distinction ~ 10

Visual Variable: Colour

✓	selective	  
✓	associative	  
≠	quantitative	
≠	order	
✓	length	 <ul style="list-style-type: none"> theoretically infinite but practically limited association and selection ~ < 7 and distinction ~ 10



Mackinlay, APT (A Presentation Tool), 1986

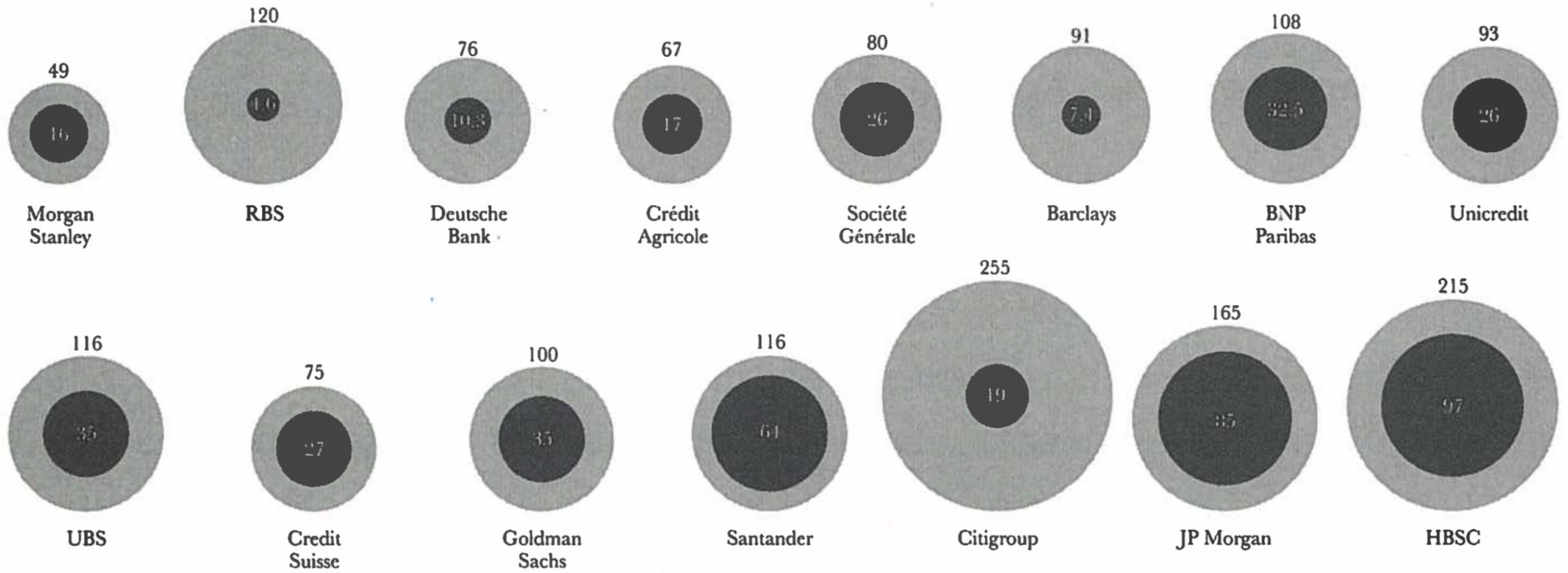
Market Capitalization of the World's Biggest Banks

In billions of dollars

January 2007

January 2009

Source: Bloomberg

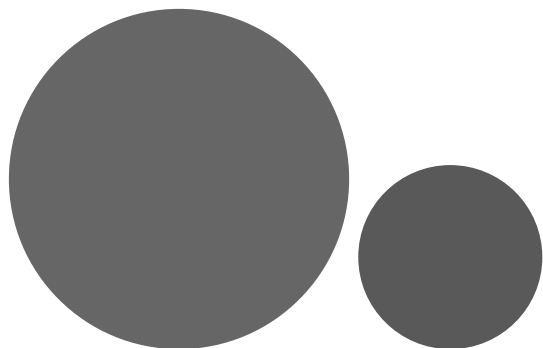
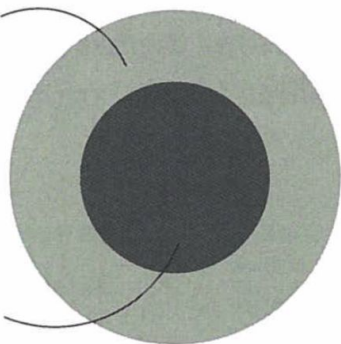


How to decode bubbles?

Billions of dollars Source: Bloomberg

January 2007

January 2009



a)



b)



c)

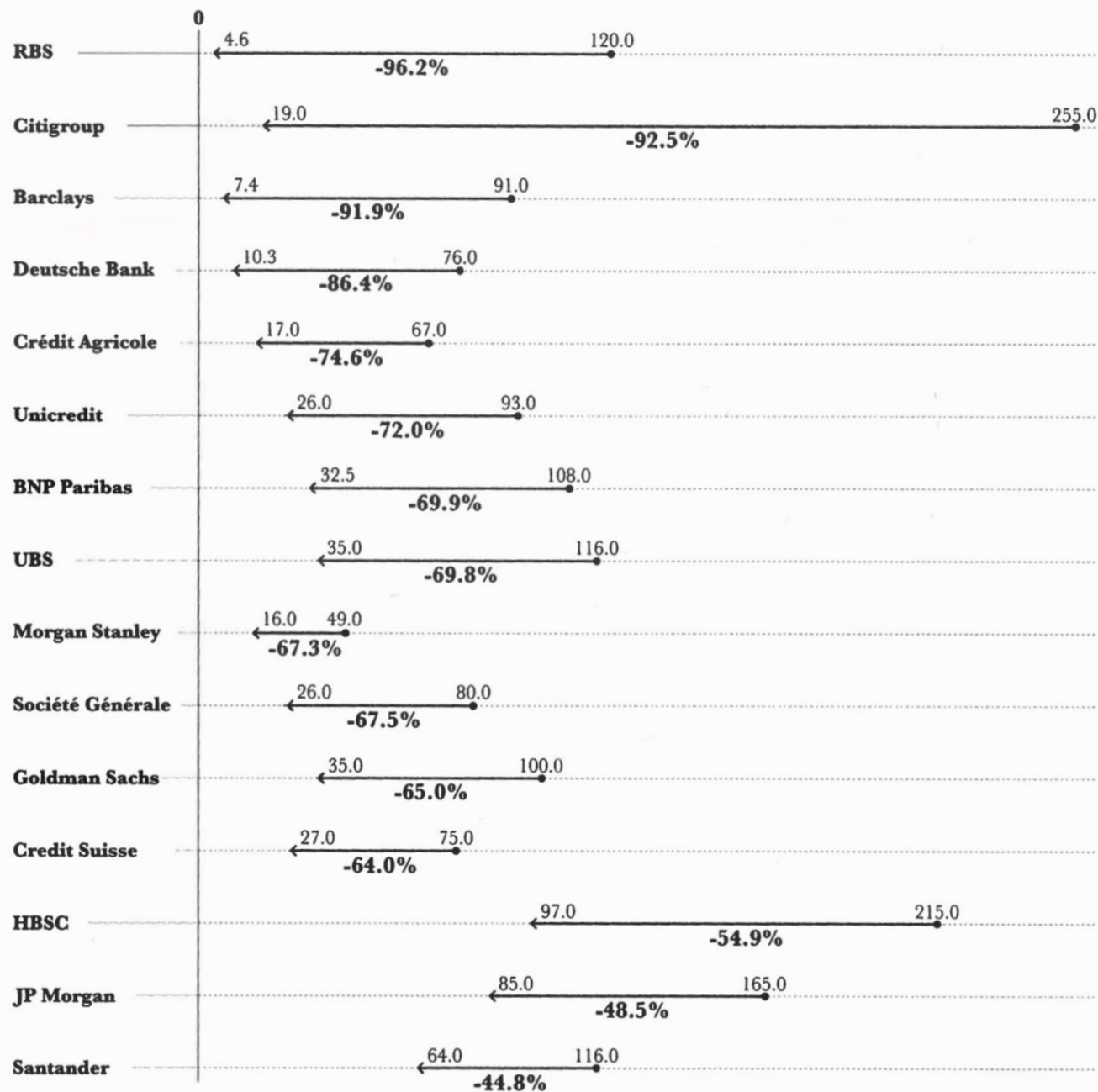


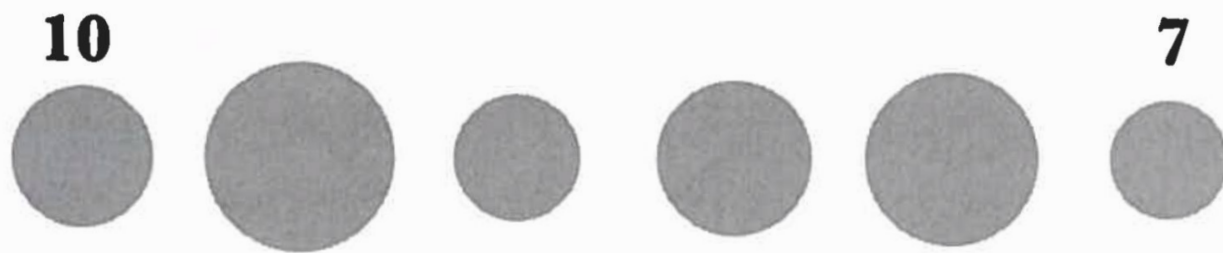
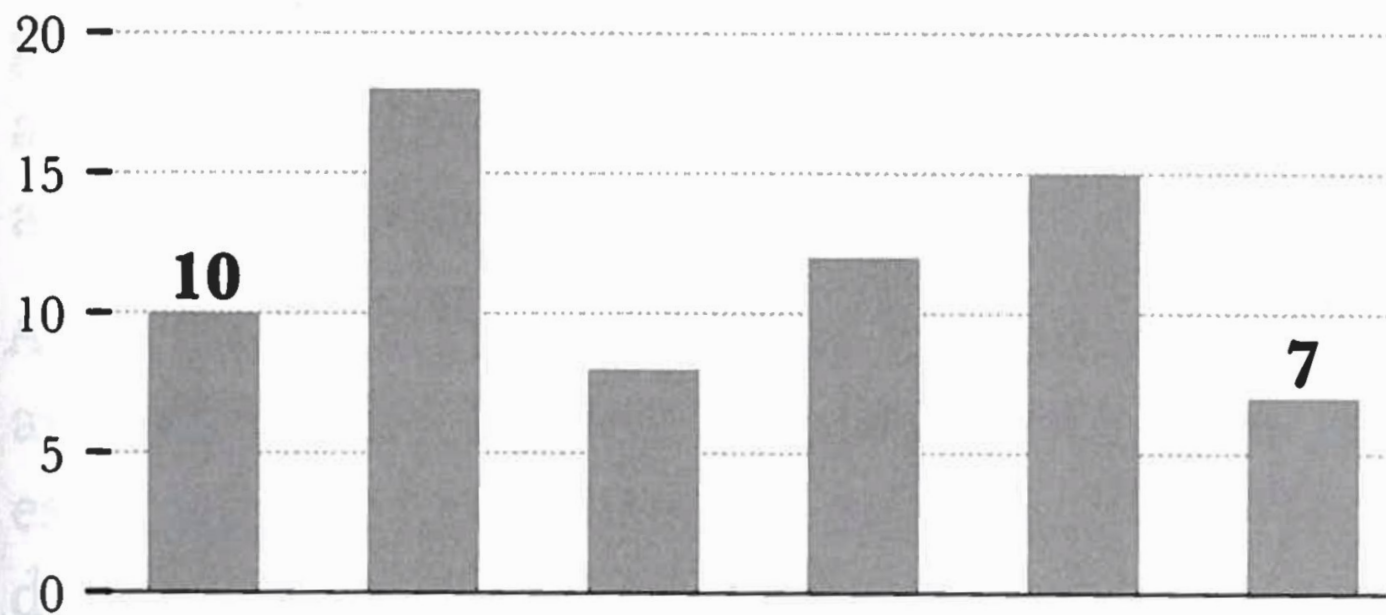
Market Capitalization of the World's Biggest Banks

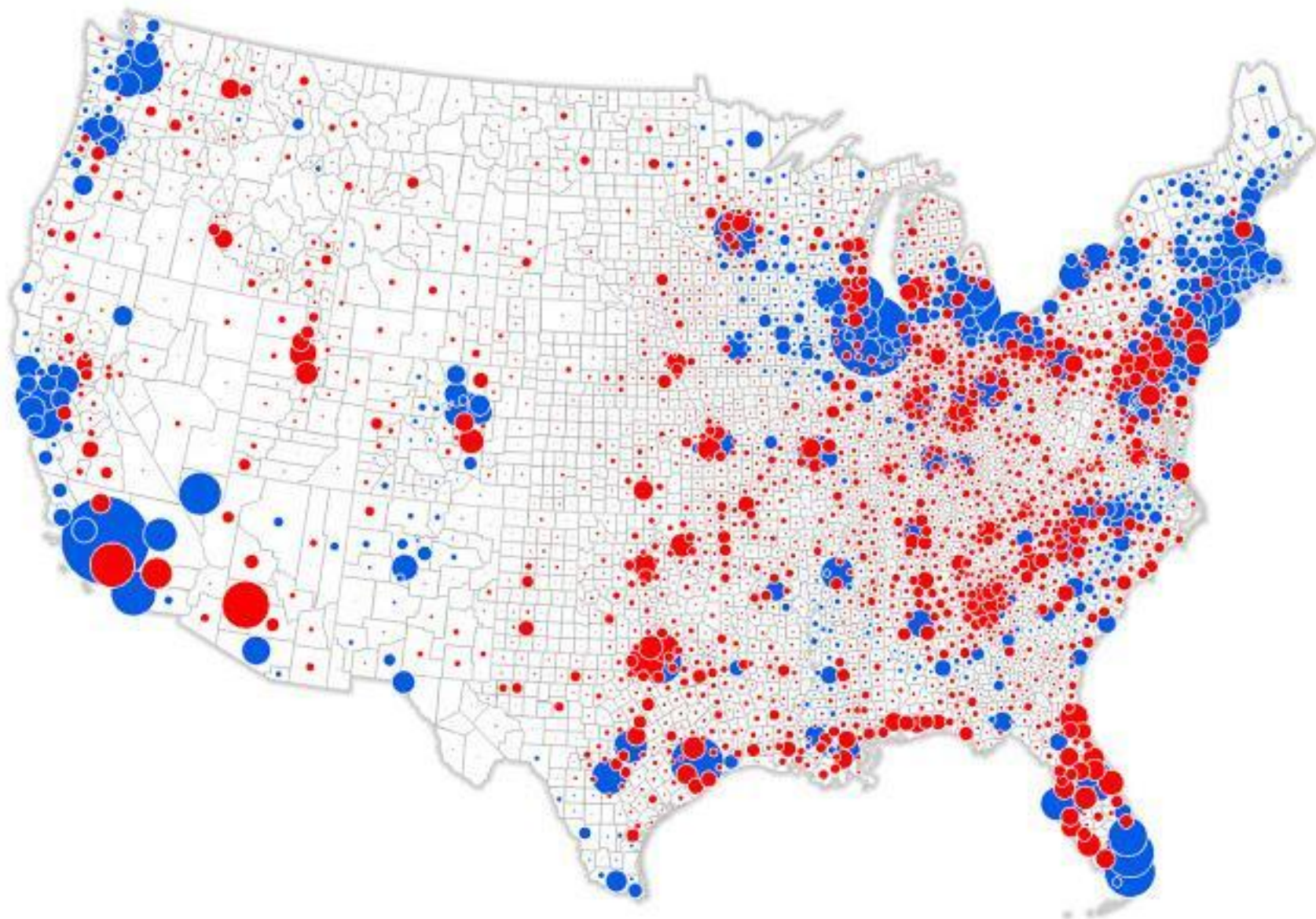
Billions of dollars

January 2009 ← January 2007

Source: Bloomberg







Recommended Reading

- Tamara Munzner: **Visualization Analysis and Design**
 - Chapter 2. What: Data Abstraction
 - Chapter 5. Marks and Channels
- Alberto Cairo: **The Functional Art**
 - Chapter 1: Why Visualize: From Information to Wisdom
 - Chapter 2: Forms and Functions: Visualization as Technology
- Andy Kirk: **Data Visualization**
 - Chapter 1: Defining Data Visualization
- Jaques Bertin: **Semiology of Graphics**
 - II.C: The retinal variables
- Cleveland & McGill: **Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods**, 1984