

Visualization

DATA 201: Thinking With Data Winter 2022

Jonathan Hudson, Ph.D
Instructor
Department of Computer Science
University of Calgary

Tuesday, March 22, 2022






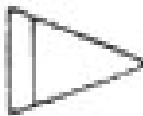





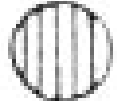













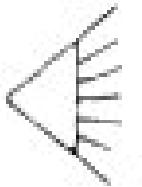
What is visualization?

Scientific Visualization
Information Visualization
Data Visualization

Mesopotamian Clay Tokens 5500 BCE

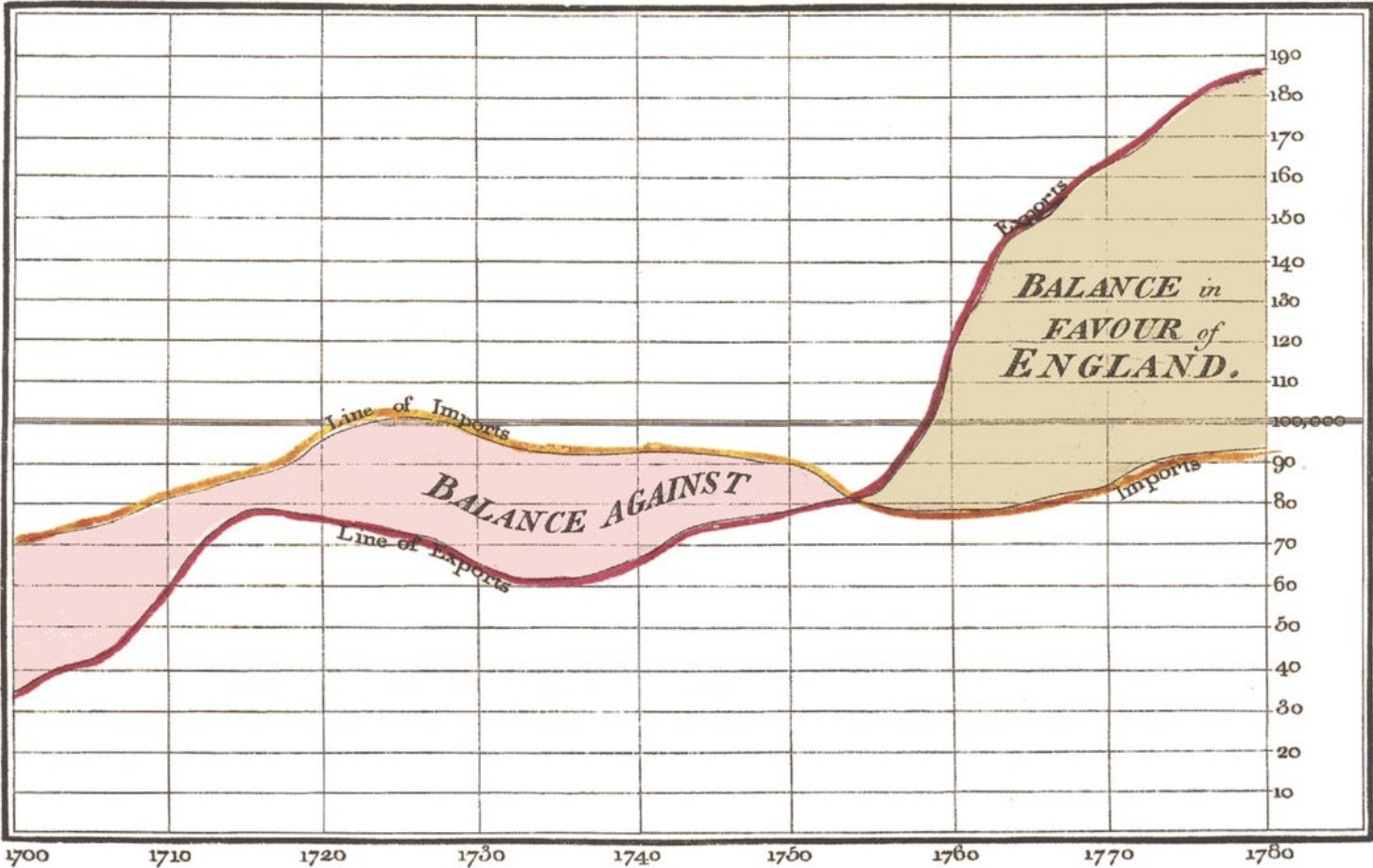


Size -> Quantity
 Shape -> Type

Tokens	Sumerian Pictographs	Tokens	Sumerian Pictographs	Tokens	Sumerian Pictographs
	 Numeral 1		 Bread		 Seat
	 Numeral 10		 Wool		 Place
	 Numeral 600		 Sheep		 Legal Decision
	 Numeral 36000		 Metal		 Cloth

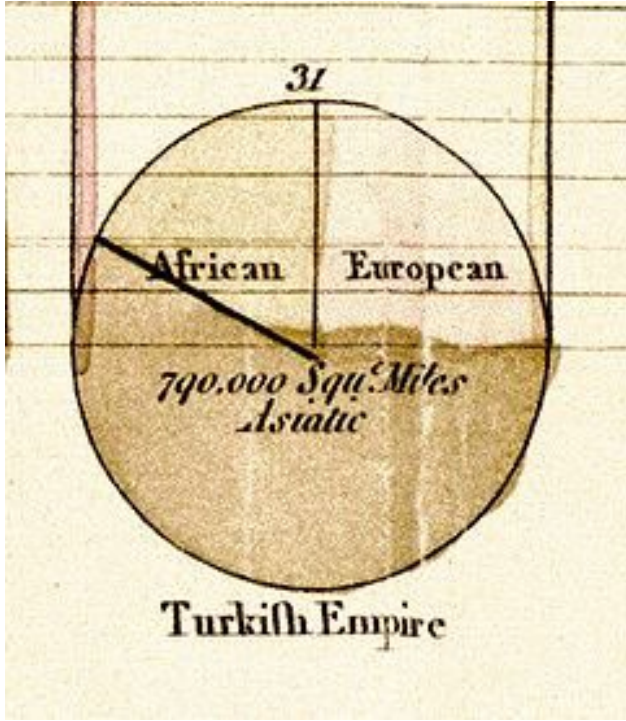
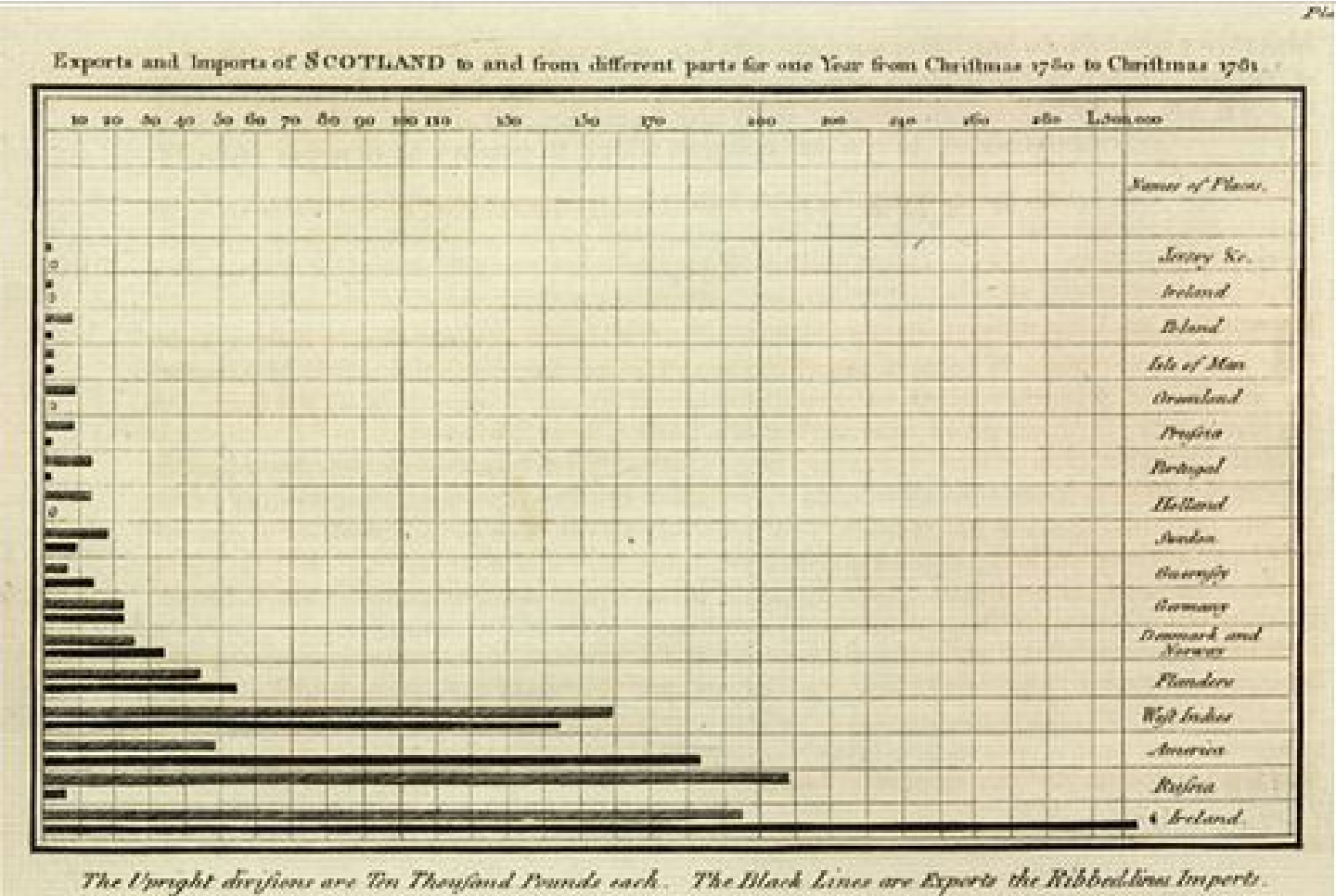
William Playfair 1786

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



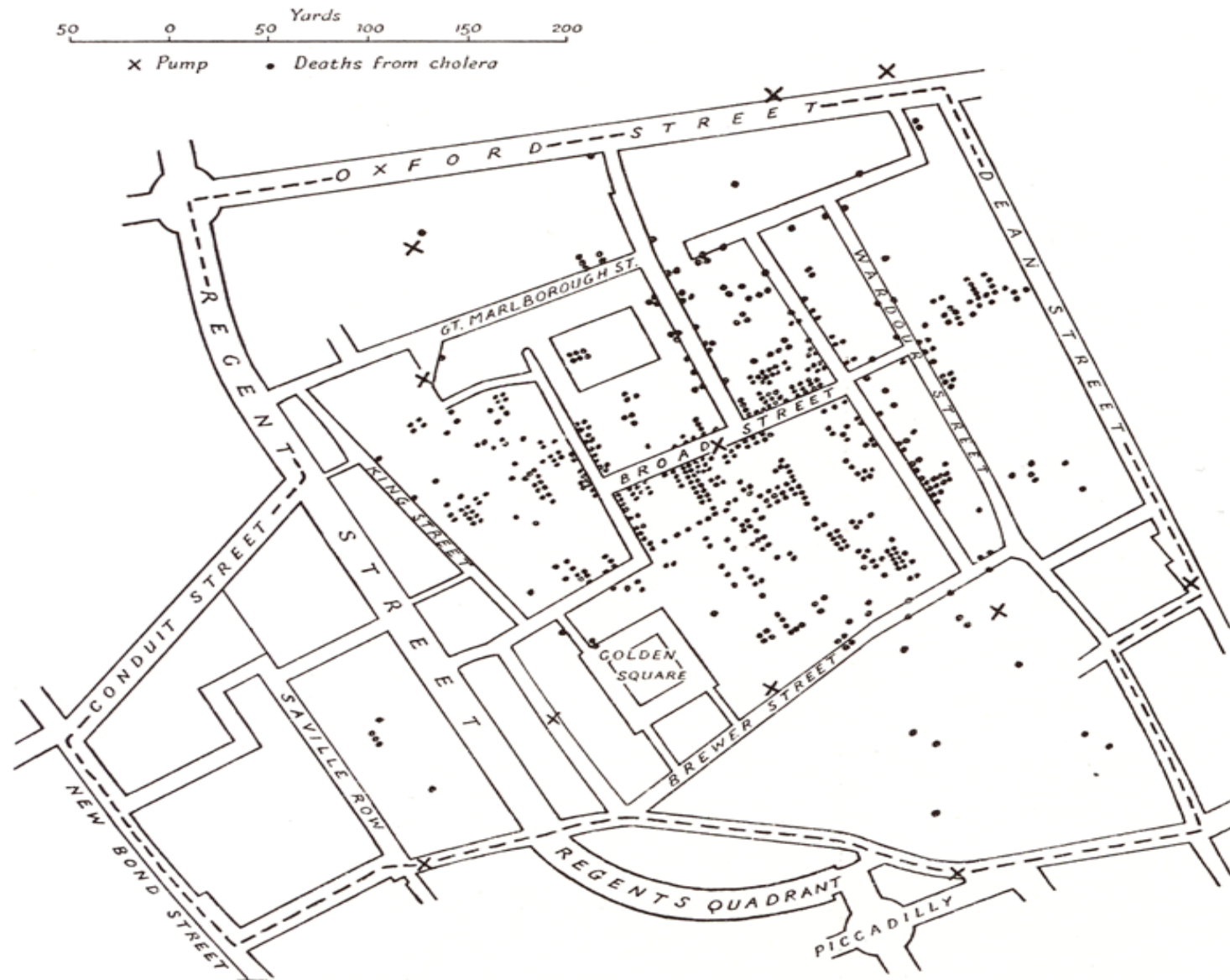
The Bottom line is divided into Years, the Right hand line into L10,000 each.
 Published as the Act directs, 1st May 1786, by W^m Playfair
 Neale sculpt 392, Strand, London.

William Playfair - Founder of graphical methods of statistics, invented numerous common diagram types



This Photo by Unknown Author is licensed under [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)

John Snow 1854



the position of each cholera case in London [from Tufte 83]

John Snow – dot chart

- Cholera in 1800s
- ‘miasma theory’ -> from ‘bad air’
- Snow -> role of water supply
- Dot map to show cases around water pump
- ‘Founding event of science of epidemiology’
- (pump dug 3 feet from old cesspit)

Jon Snow



Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

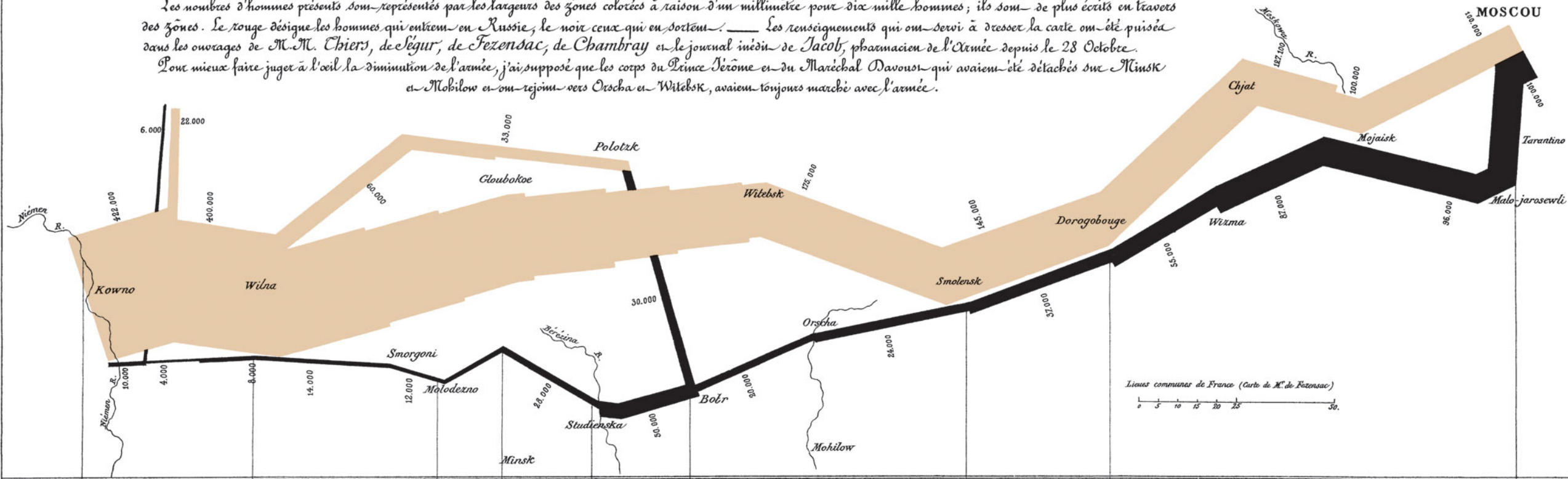
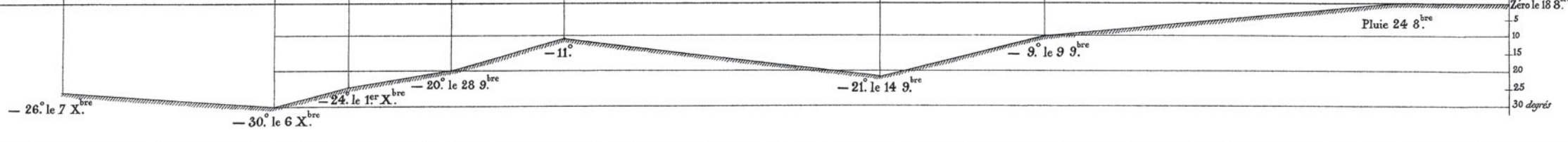


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.



Autog. par Regnier, 8. Par. 5^{to} Marie 5^{to} G^{to} à Paris.

Imp. Lith. Regnier et Douvrot.

This Photo by Unknown Author is licensed under [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)

Charles Minard

- Napoleon's losses during Russian campaign in 1812
- Six types of data
- Troop count, distance traveled, temperature, latitude, longitude, direction of travel, location relative to dates of events
- Later this type of diagram -> Sankey diagram

**Everytime a foreign power
tried to invade Russia in winter**



COVID

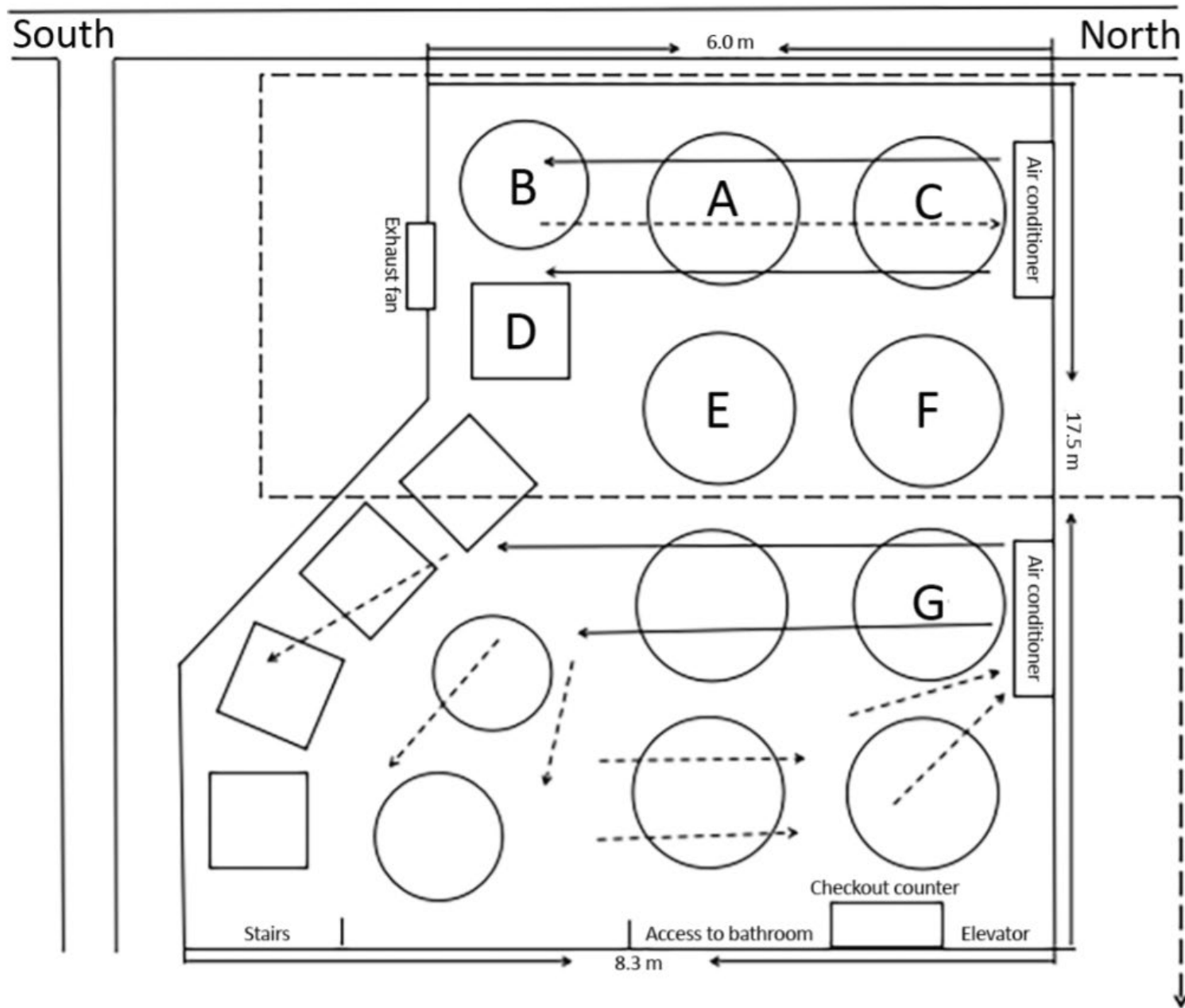


Figure. Sketch showing arrangement of restaurant tables and air conditioning airflow at site of outbreak of 2019 novel coronavirus disease, Guangzhou, China, 2020. Red circles indicate seating of future case-patients; yellow-filled red circle indicates index case-patient.

COVID

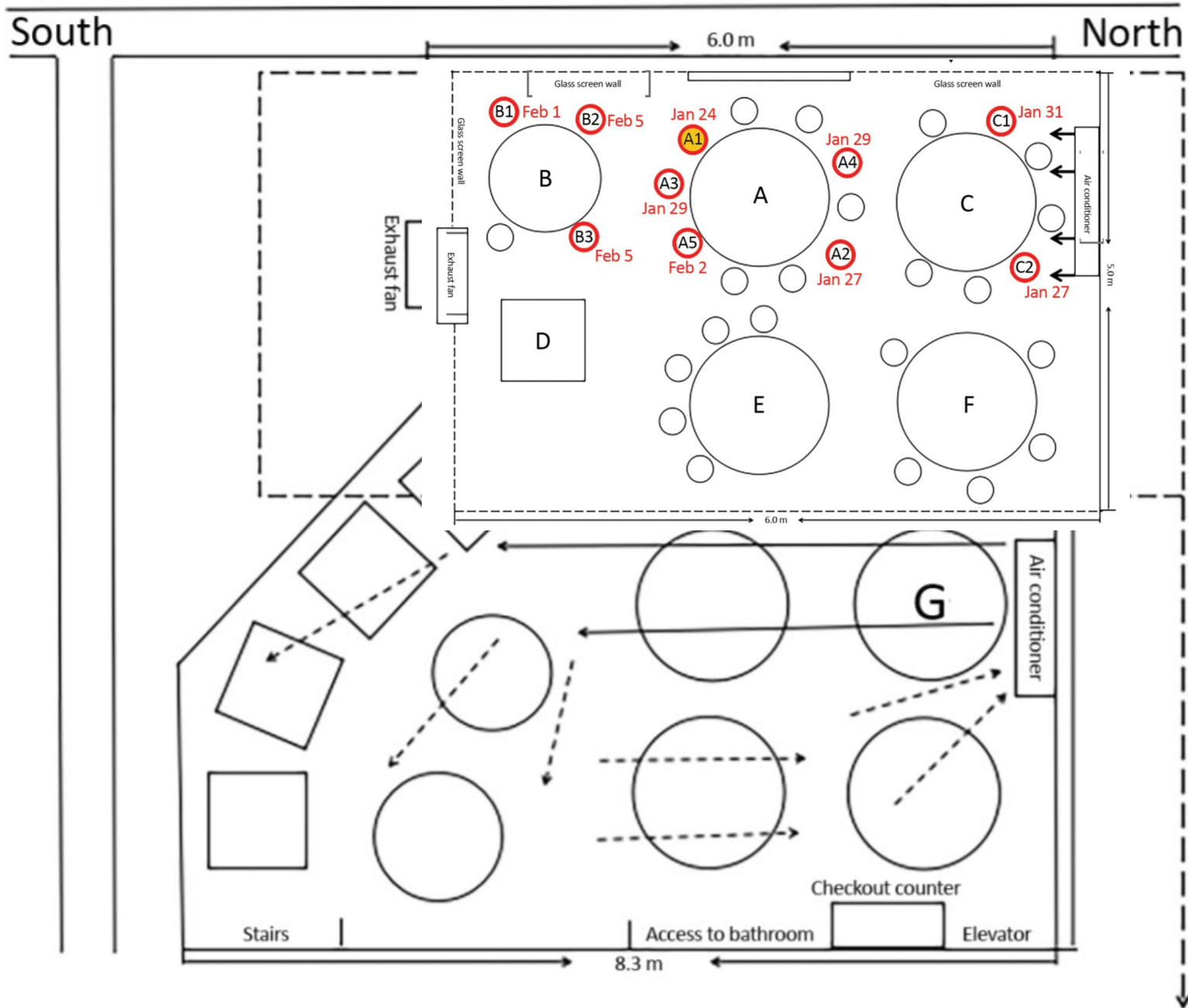
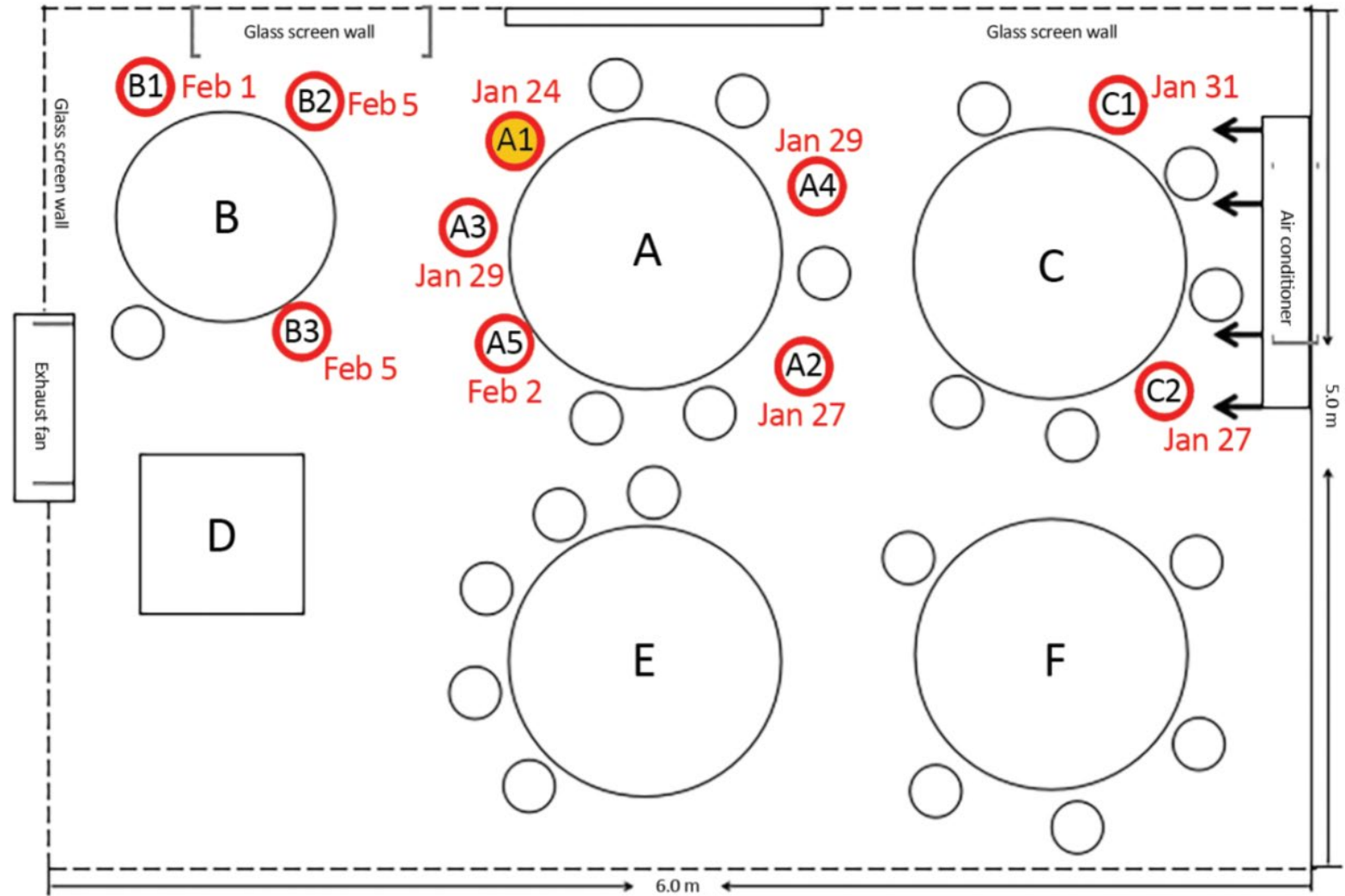
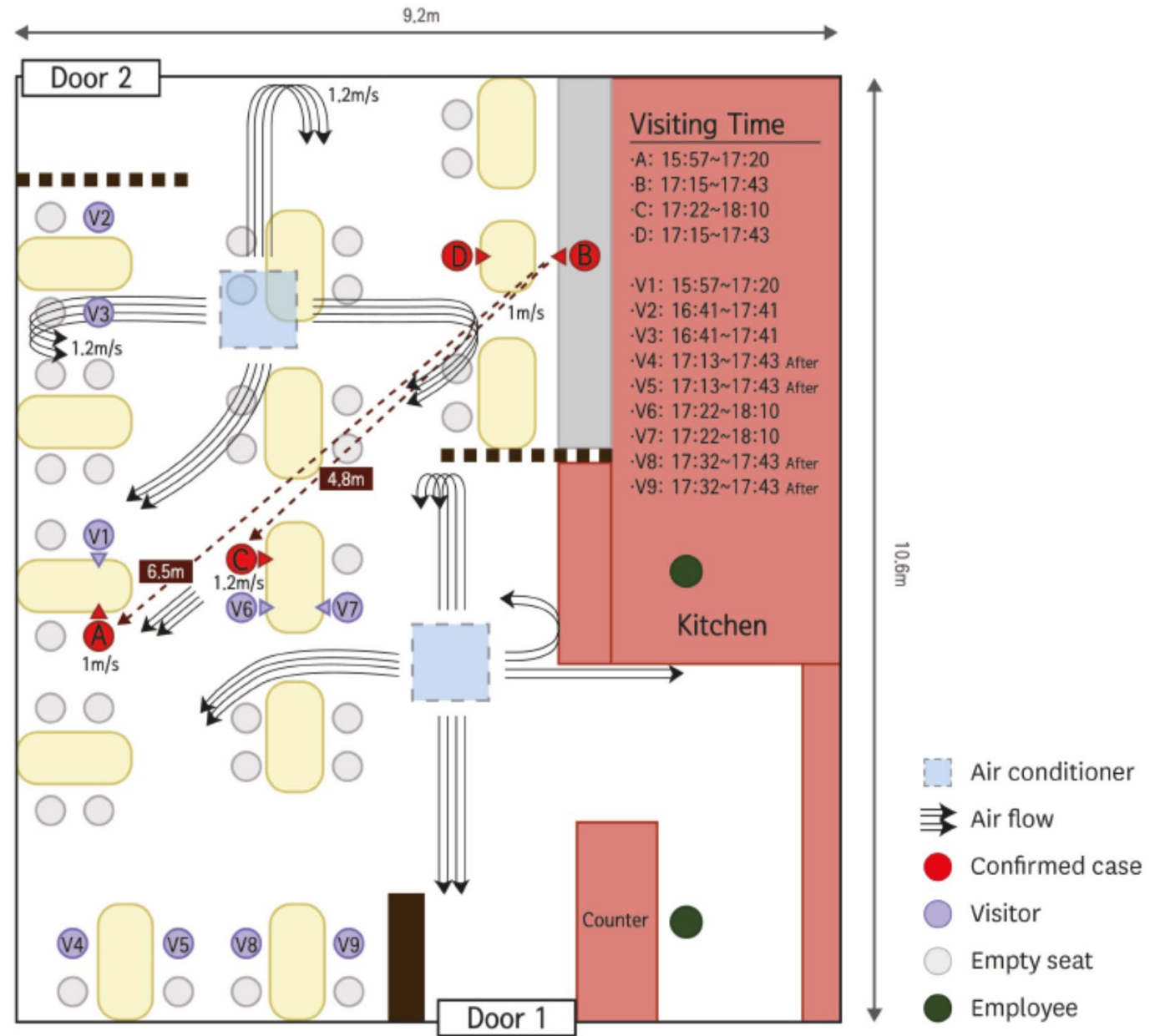


Figure. Sketch showing arrangement of restaurant tables and air conditioning airflow at site of outbreak of 2019 novel coronavirus disease, Guangzhou, China, 2020. Red circles indicate seating of future case-patients; yellow-filled red circle indicates index case-patient.

COVID



6 Feet Not Enough



COVID

- <https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/>

Visual Language

Encode

Communicates

Data =====> Images/Signs =====> Information

Levels of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

Levels of Measurement

- Nominal — categories, labels.
- Ordinal
- Interval
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal
- Interval
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
- Interval
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- Interval
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- Interval — degree of difference, arbitrary origin or arbitrary 0.
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- Interval — degree of difference, arbitrary origin or arbitrary 0.
 - E.g., temperature in C° or F°
- Ratio

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- Interval — degree of difference, arbitrary origin or arbitrary 0.
 - E.g., temperature in C° or F°
- Ratio — unique (non-arbitrary) 0 value. Zero indicates the absence of the quantity.

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- Interval — degree of difference, arbitrary origin or arbitrary 0.
 - E.g., temperature in C° or F°
- Ratio — unique (non-arbitrary) 0 value. Zero indicates the absence of the quantity.
 - E.g., length

Levels of Measurement

- Nominal — categories, labels.
 - E.g., fruits: apples, oranges, bananas, etc.
- Ordinal — meaningful order.
 - E.g., medals: gold, silver, bronze
- **Quantitative Interval** — degree of difference, arbitrary origin or arbitrary 0.
 - E.g., temperature in C° or F°
- **Quantitative Ratio** — unique (non-arbitrary) 0 value. Zero indicates the absence of the quantity.
 - E.g., length

Levels of Measurement – Tour de France

- Nominal — categories, labels.
 - Team Sky, sprinter/climber/etc.
- Ordinal — meaningful order.
 - Podium finish (maybe quantitative but height often arbitrary on podium)
- **Quantitative Interval** — degree of difference, arbitrary origin or arbitrary 0.
 - Rank ordering. 1st to cross finish line, 2nd to cross finish line (not a quantity where 0th means anything)
- **Quantitative Ratio** — unique (non-arbitrary) 0 value. Zero indicates the absence of the quantity.
 - Time back of leader (leader at 0, second place 33s back, third place 1:33 back)

Relational Data Model

- Relation (Table)
- Tuple (Row) Attribute (Column)
- Schema (Blueprint / table structure)
Database (A collection of relation)

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Blood pressure study (4 treatments, 6 months)

Relational Data Model

- Dimensions
 - discrete variables
 - e.g., categories, names
- Measures
 - can be aggregated usually continuous
 - e.g., weight, height

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Blood pressure study (4 treatments, 6 months)

Keys

- **Primary key**

- A column
- Each row value unique in this table of data
- Each record uniquely connected to this
- Used by program to identify row
- Only one

- **Secondary key**

- A column
- Each row value unique in this table of data
- Each record uniquely connected to this
- Not used by program to identify row
- Can be zero, one, or more

No	ID	Name
1	3012143	Jon
2	3002243	Jon
3	3102143	Jonathan
4	3002144	John
5	3002121	Dr. J
6	3006143	John
7	3802142	Jonathan
8	3402143	Jon
9	3003243	Johnathan

Keys

- Foreign Key
 - A primary key in another table

No	ID	Name
1	3012143	Jon
2	3002243	Jon
3	3102143	Jonathan

No	ID	Course	Grade
1	3012143	DATA 201	A
2	3012143	DATA 211	A-
3	3012143	DATA 311	B+

Keys

- **Foreign Key**

- A primary key in another table

No	ID	Name
1	3012143	Jon
2	3002243	Jon
3	3102143	Jonathan

No	ID	Course	Grade
1	3012143	DATA 201	A
2	3012143	DATA 211	A-
3	3012143	DATA 311	B+

Keys

- **Foreign Key**

- A primary key in another table
- Used to join tables together
- (Note that we didn't have to store 'Jon' as a name for every single grade)

No	ID	Course	Grade	Name
1	3012143	DATA 201	A	Jon
2	3012143	DATA 211	A-	Jon
3	3012143	DATA 311	B+	Jon

Example (Census)

- Levels of Measurement (nominal, ordinal, interval, or ratio)
- Types of Attribute (dimension or measure)

- Year: 1901 – 2016 (every 5 years)
- Age: 0 – 90+
- Marital Status: Single, Married, Divorced,...
- People: # of people in group

Dimensions and Measures are important concepts in many analysis tools.

Data with different levels of measurement are best to encode in different ways.

Encoding

Marks

Marks



Points



Areas



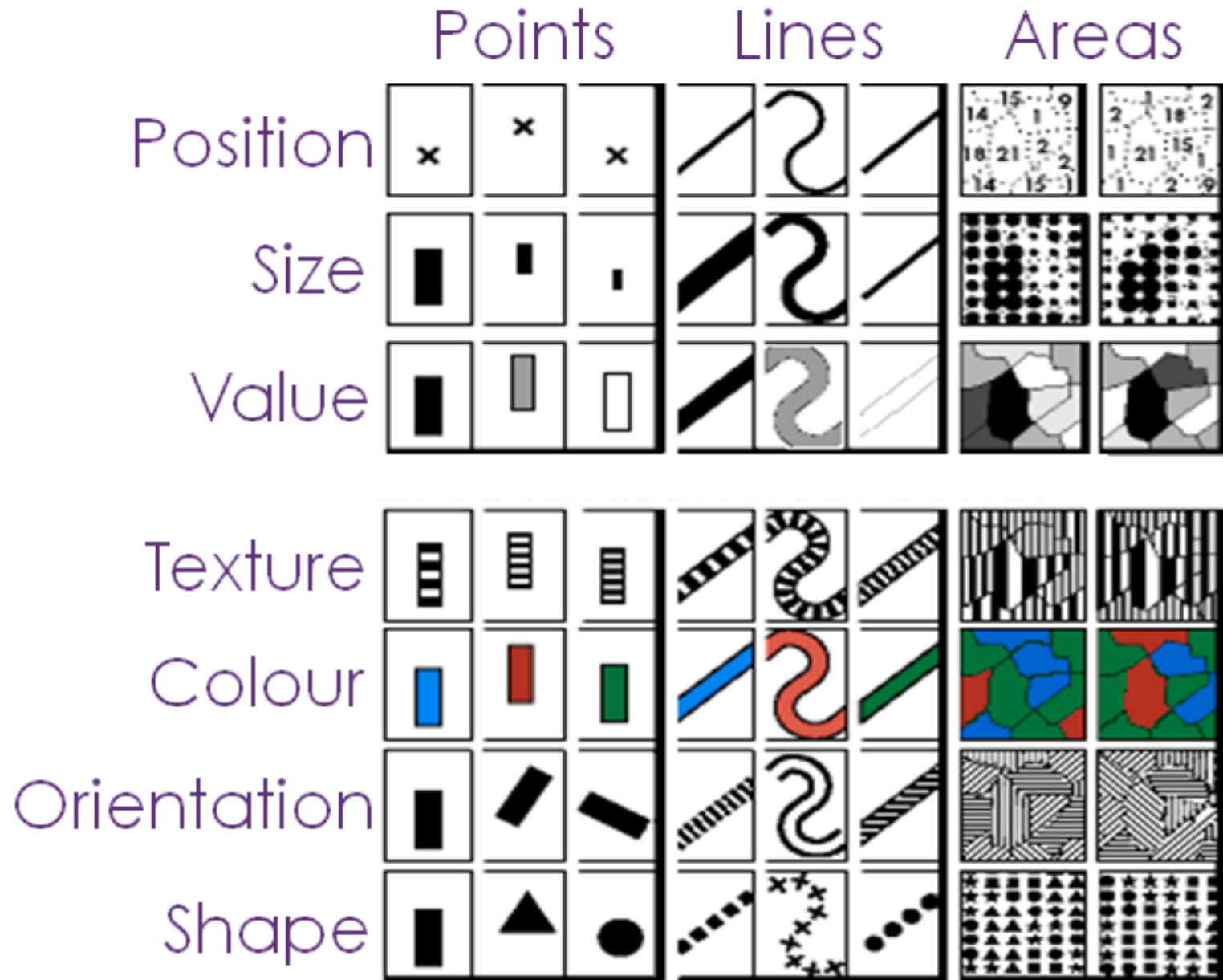
Lines

Visual Variables


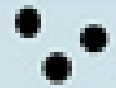






















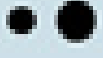


Visual Variables

- Position
- Size
- Value
- Texture
- Colour
- Orientation
- Shape

Bertin 1974












	Nominal	Ordinal	Quantitative
Position	★	★	★
Size	★	★	★
Value	★	★	★
Texture	★	★	
Colour	★		
Orientation	★		
Shape	★		

	Nominal	Ordinal	Quantitative
More Accurate	Position 	Position 	Position 
	Hue 	Density 	Length 
	Density 	Saturation 	Angle 
	Saturation 	Hue 	Slope 
	Shape 	Length 	Area 
	Length 	Angle 	Density 
	Angle 	Slope 	Saturation 
	Slope 	Area 	Hue 
Less Accurate	Area 	Shape 	Shape 

Jacques Bertin refined by Cleveland & McGill then by Card & Mackinlay

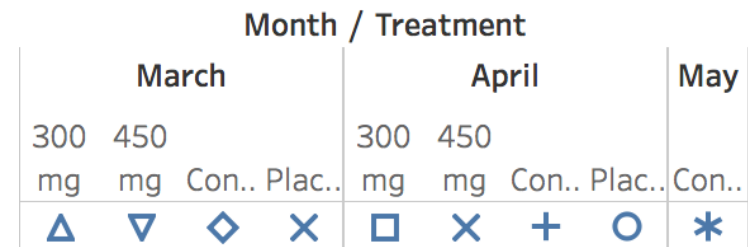
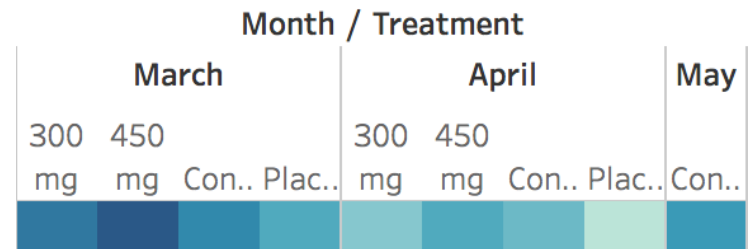
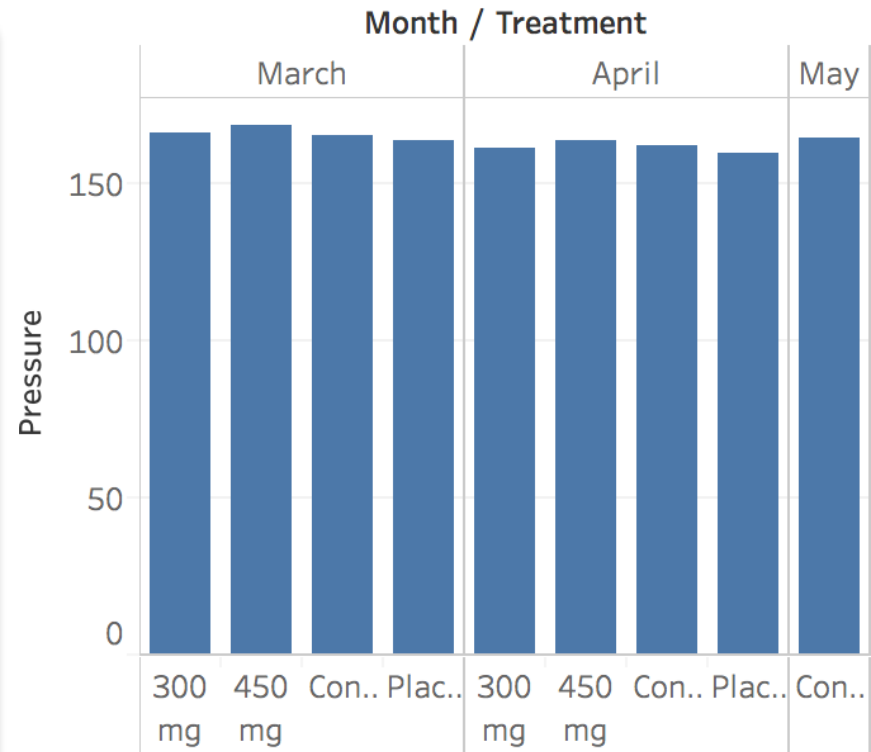
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Quantitative	
Position	
Length	
Angle	
Slope	
Area	
Density	
Saturation	
Hue	
Shape	

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164




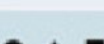
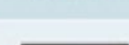
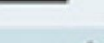

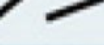
Quantitative	
Position	
Length	
Angle	
Slope	
Area	
Density	
Saturation	
Hue	
Shape	



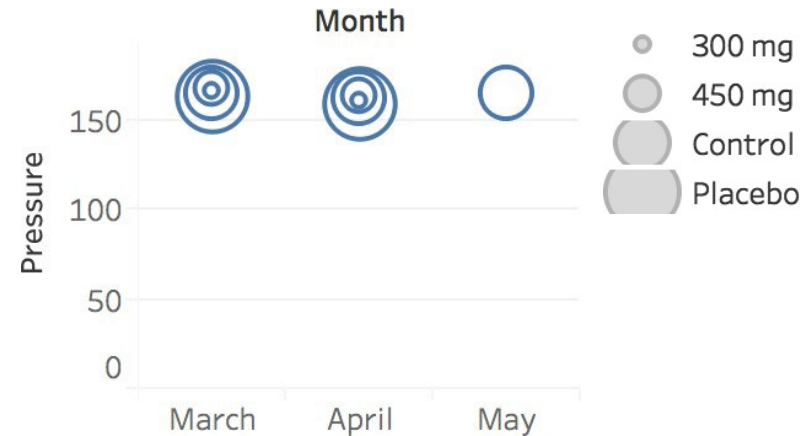
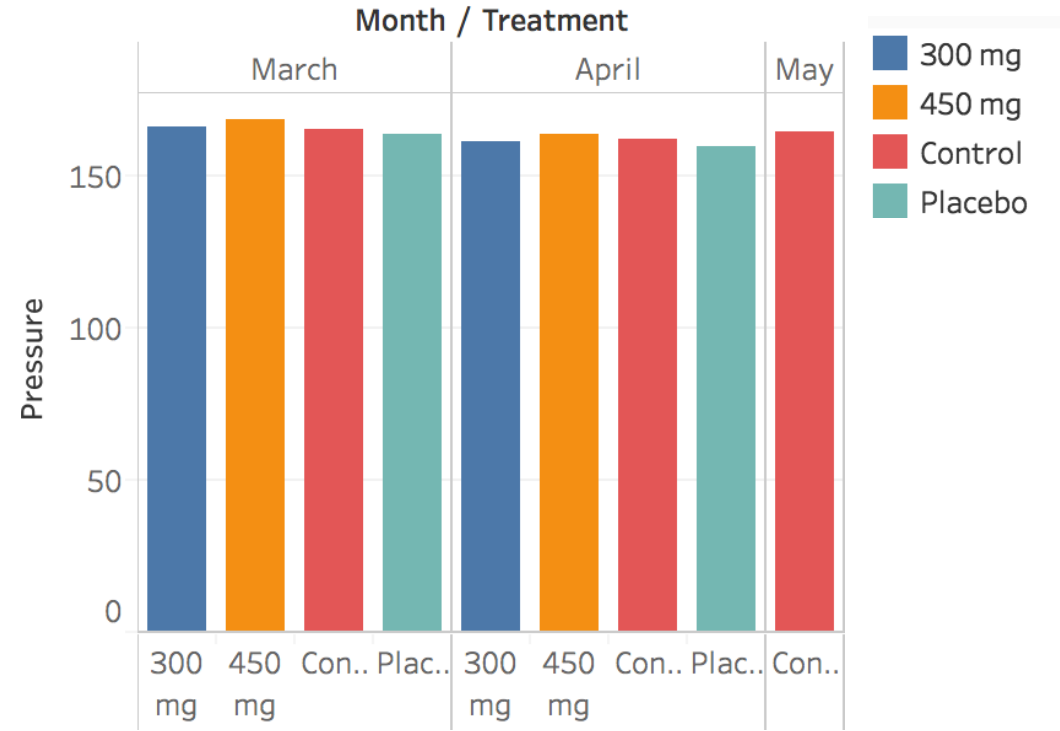
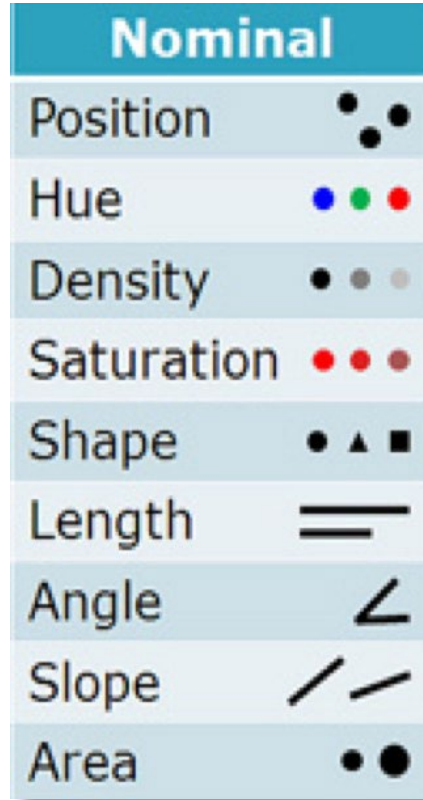
- 159
- 161
- +
- × 163
- * 164
- ◇ 165
- △ 166
- ▽ 168

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Nominal	
Position	
Hue	
Density	
Saturation	
Shape	
Length	
Angle	
Slope	
Area	

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164



More Variables with Computers

Motion



More Variables with Computers

Motion

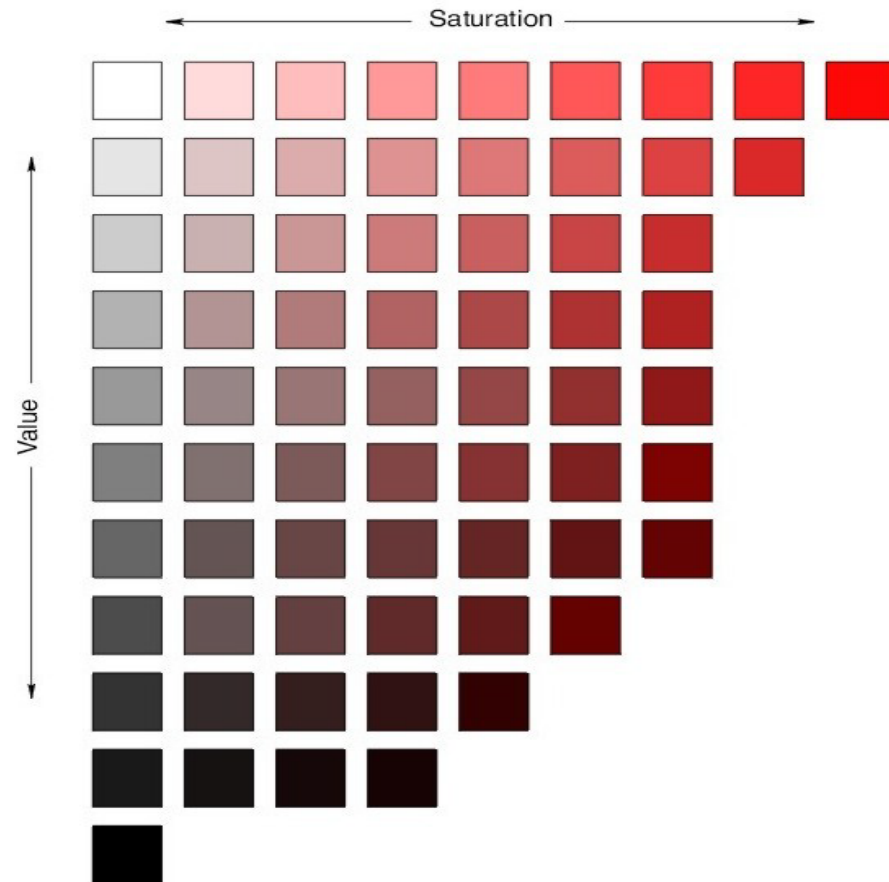


Colour

Hue (actual colour)

Value (brightness)

Saturation (intensity)



More Variables with Computers

Flicker

Frequency, rhythm



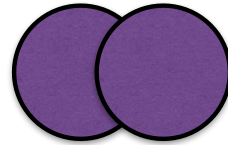
More Variables with Computers

Flicker

Frequency, rhythm



Depth



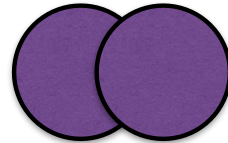
More Variables with Computers

Flicker

Frequency, rhythm

Depth

Transparency



Characteristics of Visual Variables

- Selective
- Associative
- Quantitative
- Order
- Variations / Length / Resolution

Characteristics of Visual Variables

- Selective — differentiate items from groups
- Associative
- Quantitative
- Order
- Variations / Length / Resolution

Characteristics of Visual Variables

- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative
- Order
- Variations / Length / Resolution

Characteristics of Visual Variables

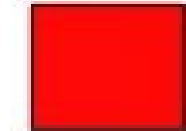
- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order
- Variations / Length / Resolution

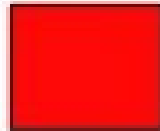
Characteristics of Visual Variables

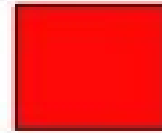
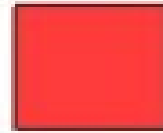
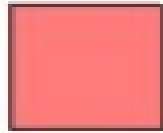
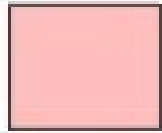
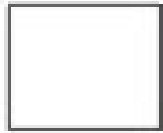
- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order — perceive an order
- Variations / Length / Resolution

Characteristics of Visual Variables

- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order — perceive an order
- Variations / Length / Resolution — distinguishable variations (How many variations in the visual variable are distinctions recognizable?)









That's too many slices!

Position

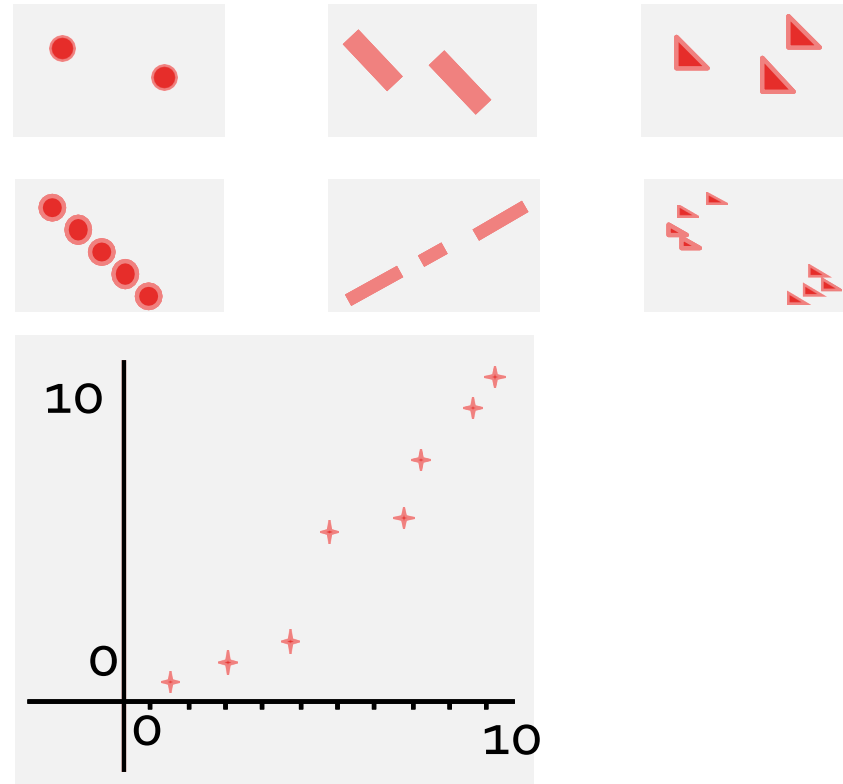
Selective

Associative

Quantitative

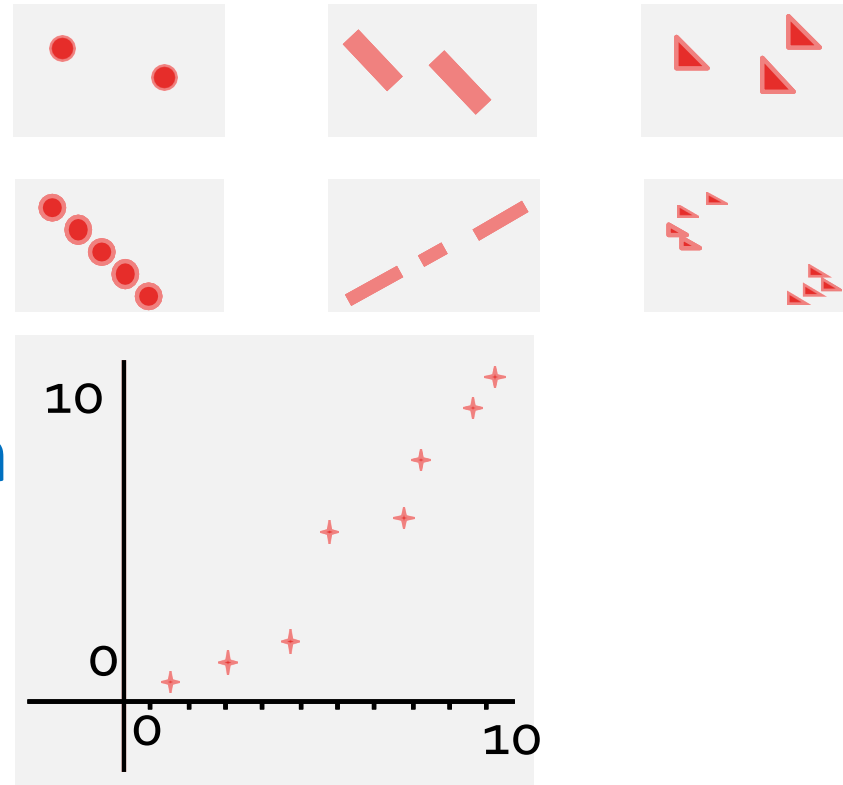
Order

Variations / Length / Resolution



Position

- + Selective
- + Associative
- + Quantitative
- + Order
- + Variations / Length / Resolution



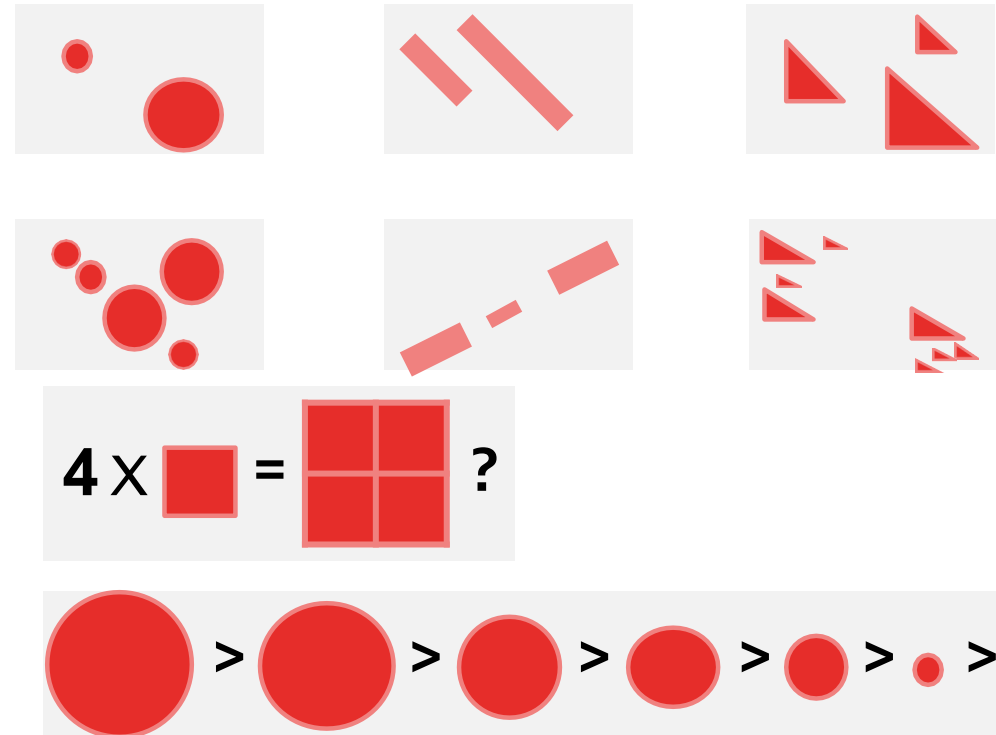
Size

Selective
Associative

Quantitative

Order

Variations / Length / Resolution



Size

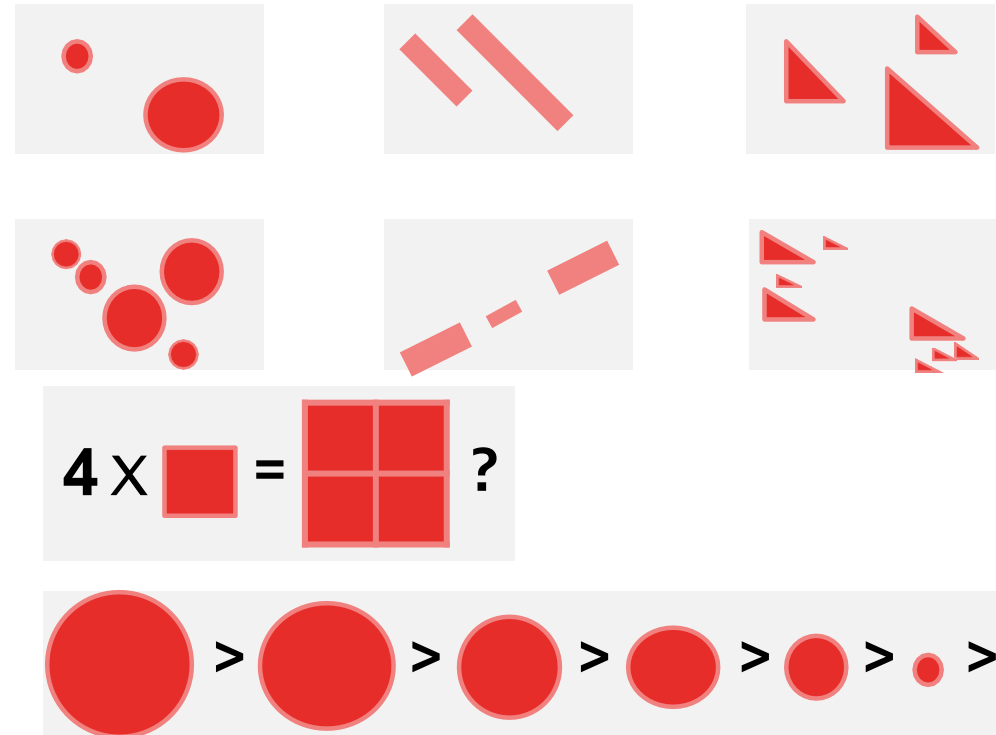
+ Selective

+ Associative

+/- Quantitative

+ Order

+ Variations / Length / Resolution (theory inf) practically limited



Shape



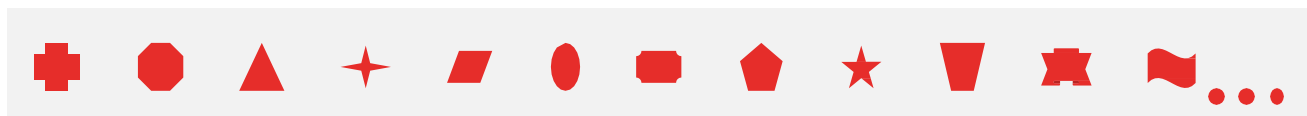
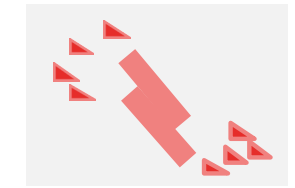
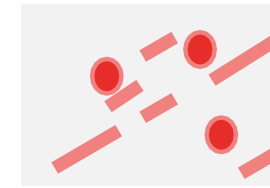
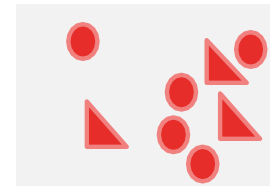
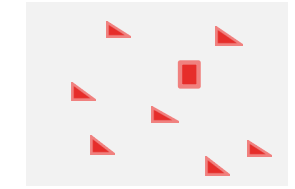
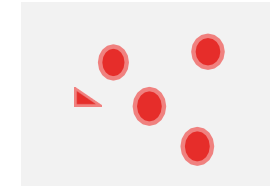
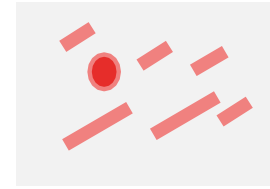
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Shape

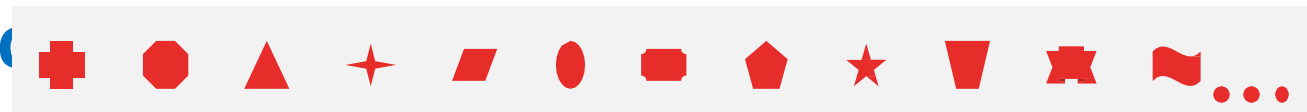
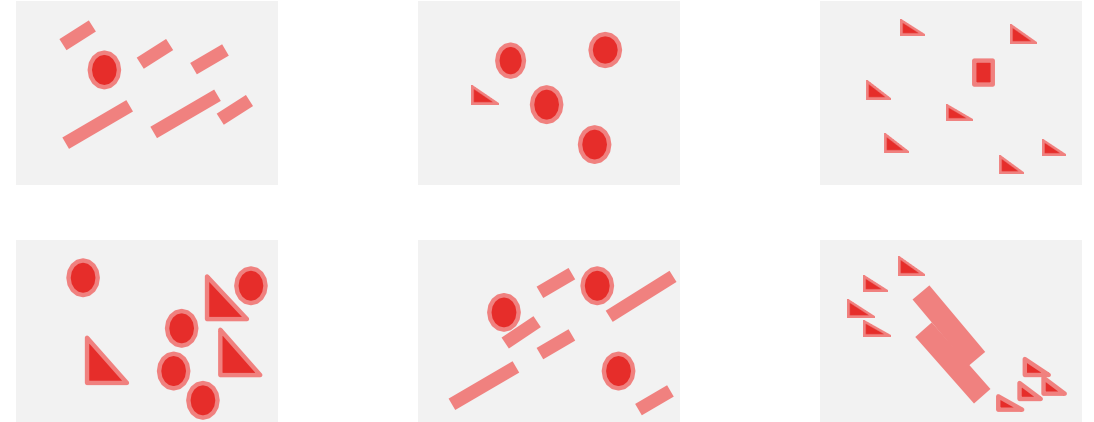
+/- Selective

+/- Associative

- Quantitative

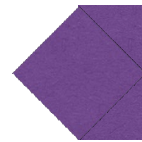
- Order

+ Variations / Length / Resolution

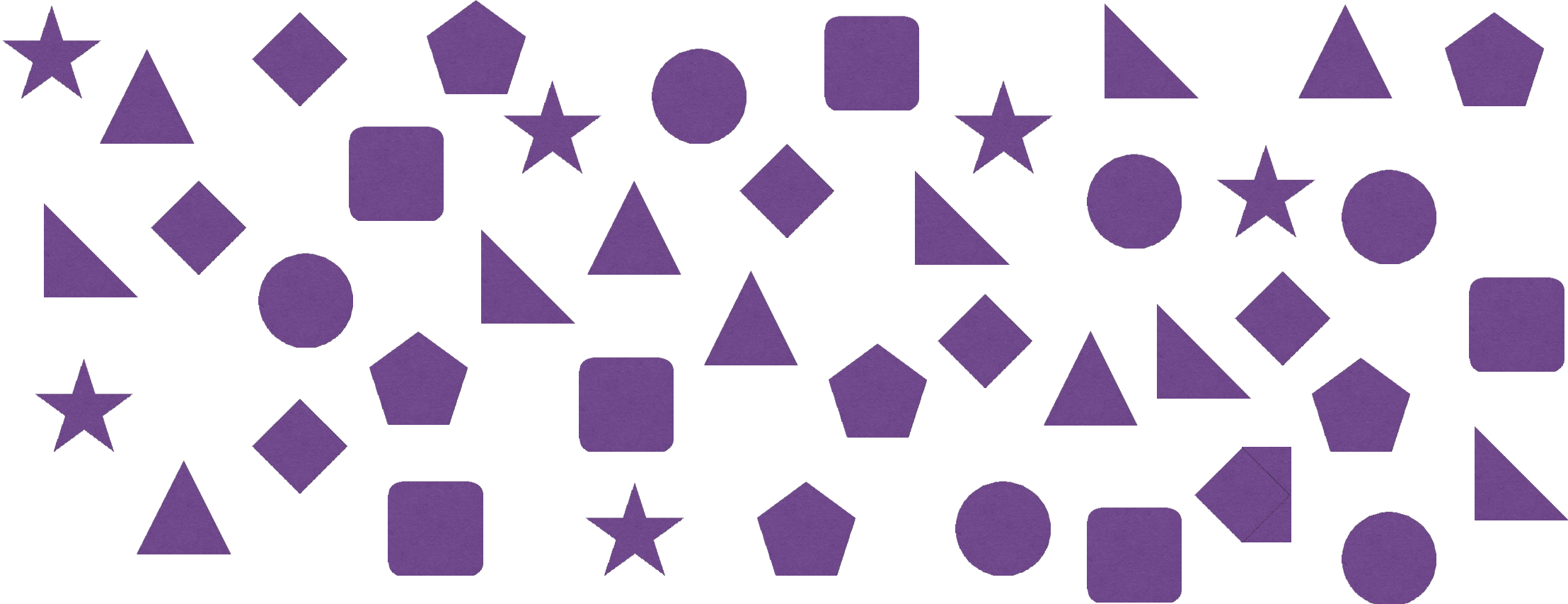


Shape

Find



Shape



Value



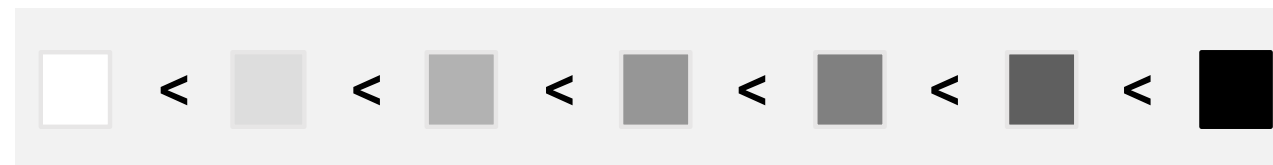
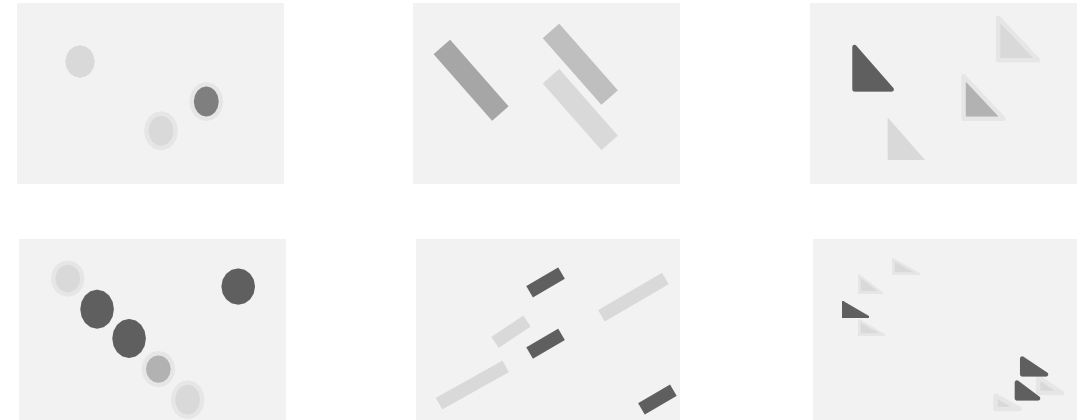
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Value

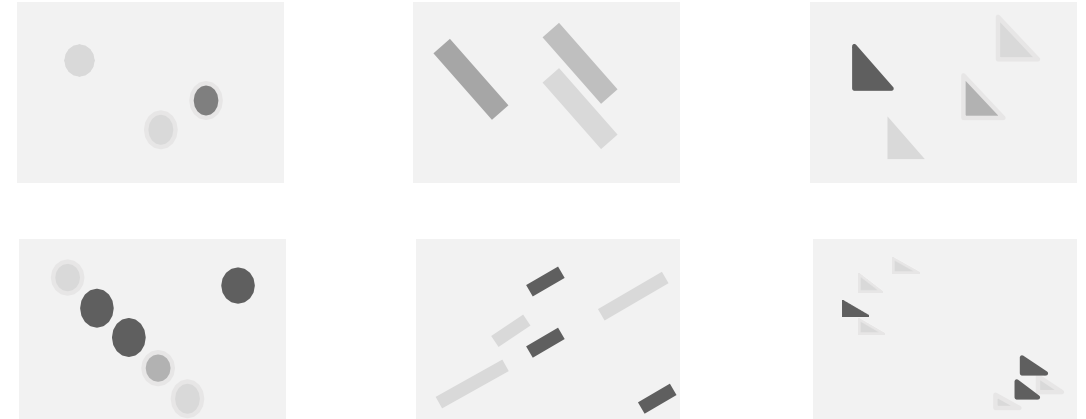
+ Selective

+ Associative

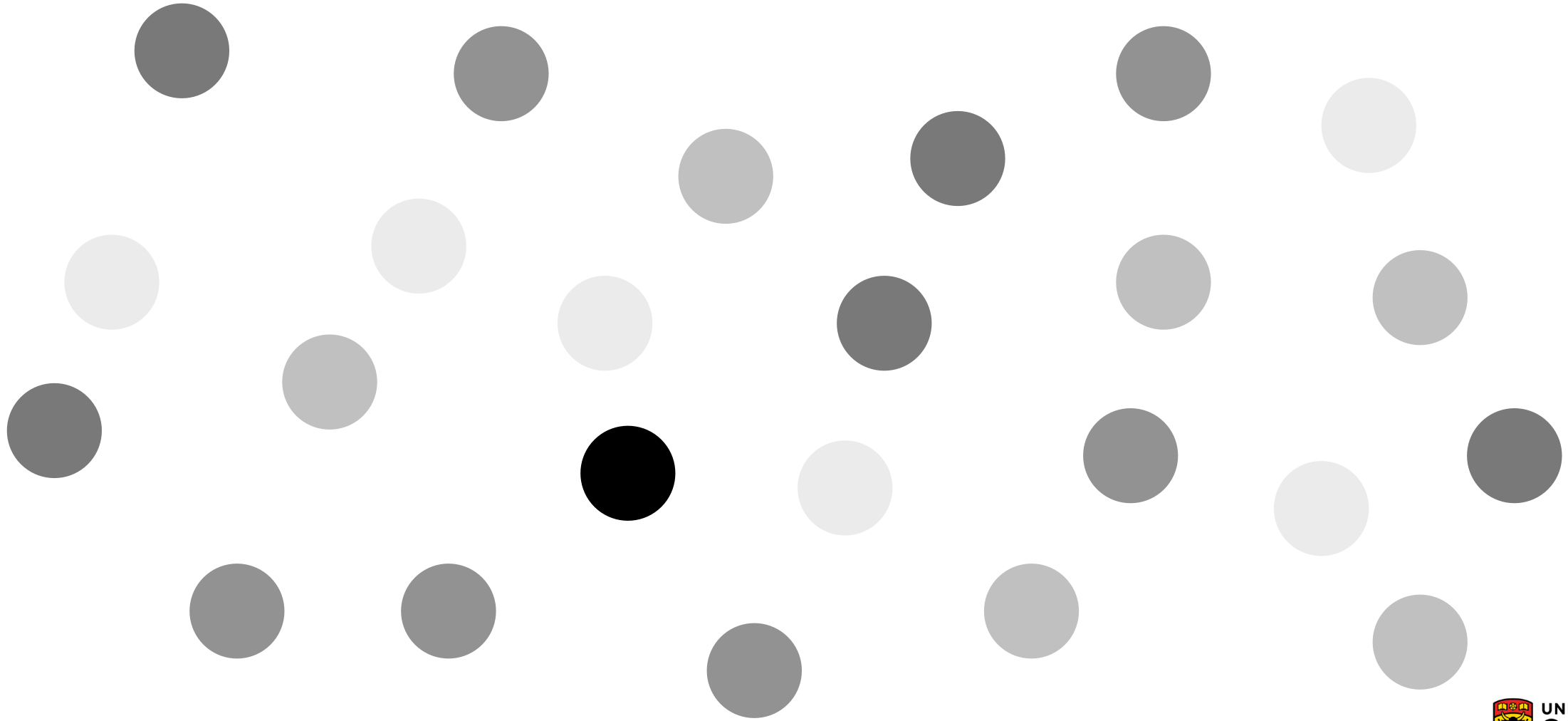
- Quantitative

+ Order

+ Variations / Length / Resolution (theory inf., practically limited)

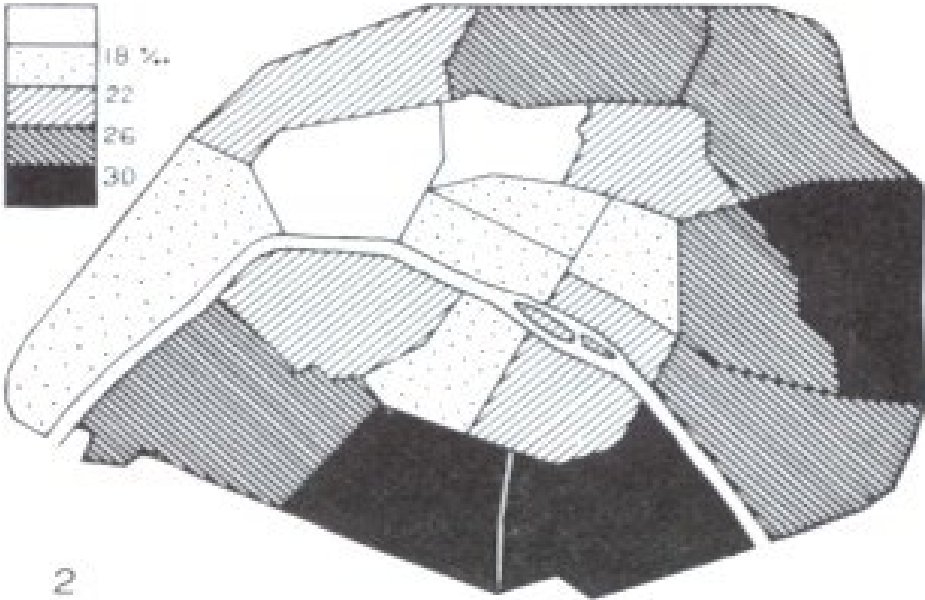
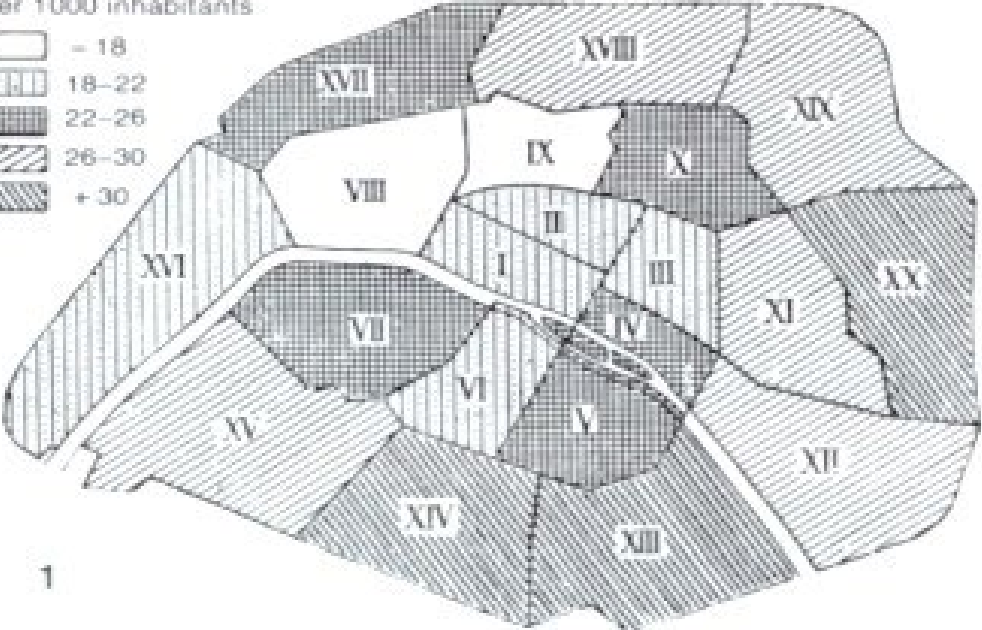


Value



Value

ANNUAL DEATHS
per 1000 inhabitants



Colour

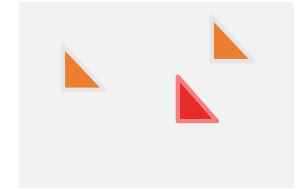
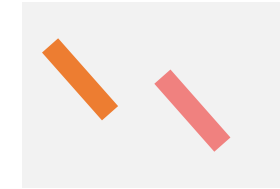
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Colour

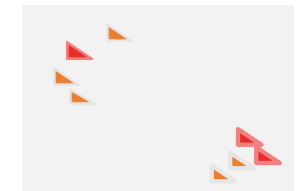
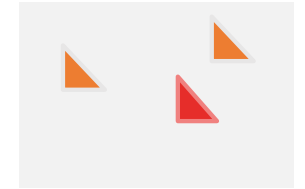
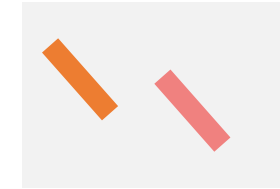
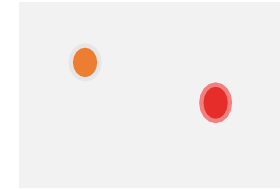
+ Selective

+ Associative

- Quantitative

- Order

+ Variations / Length / Resolution (theory inf, practically limited)



Colour-blind



NORMAL VISION



DEUTERANOMALY



PROTANOPIA

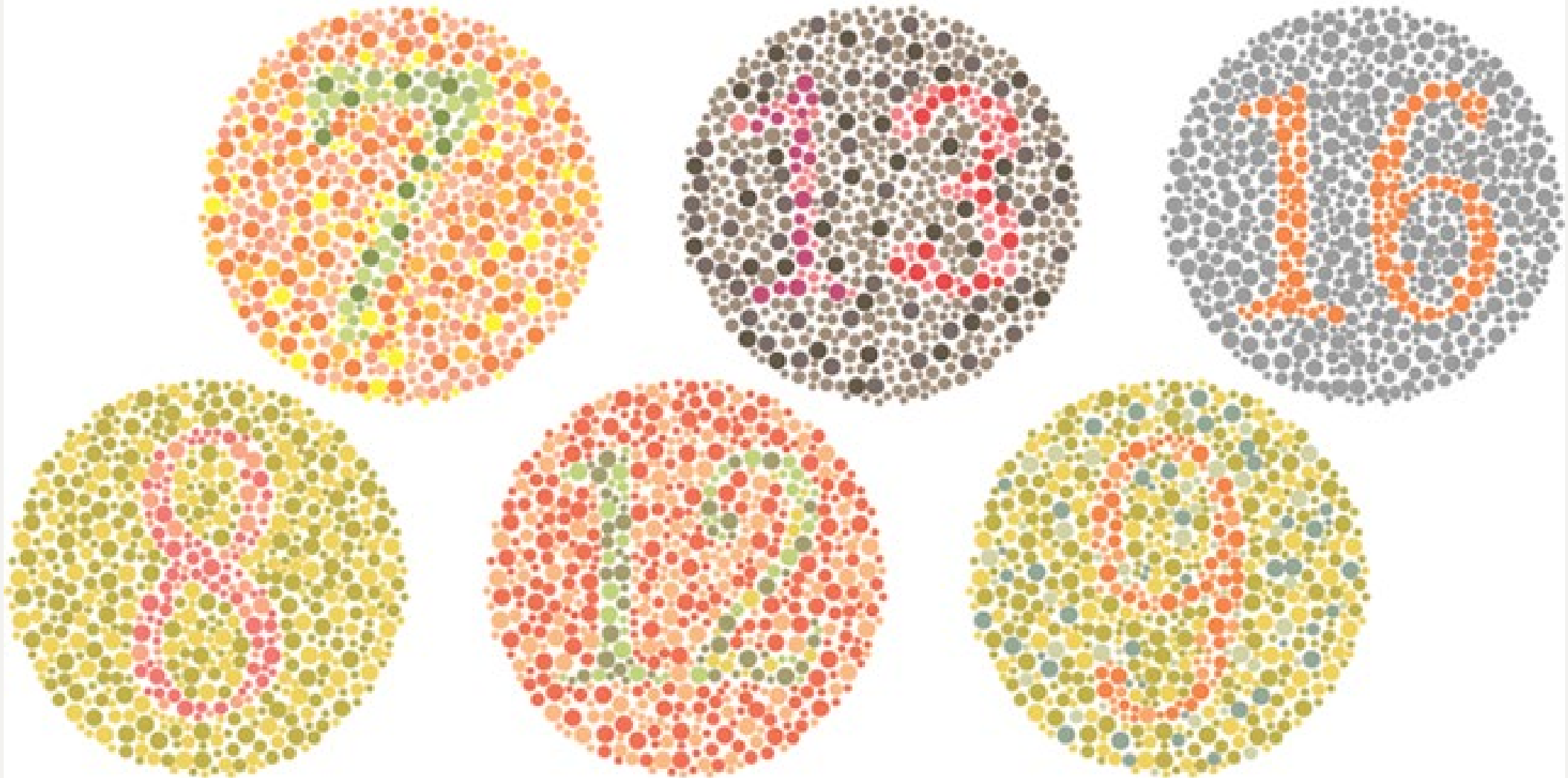


TRITANOPIA



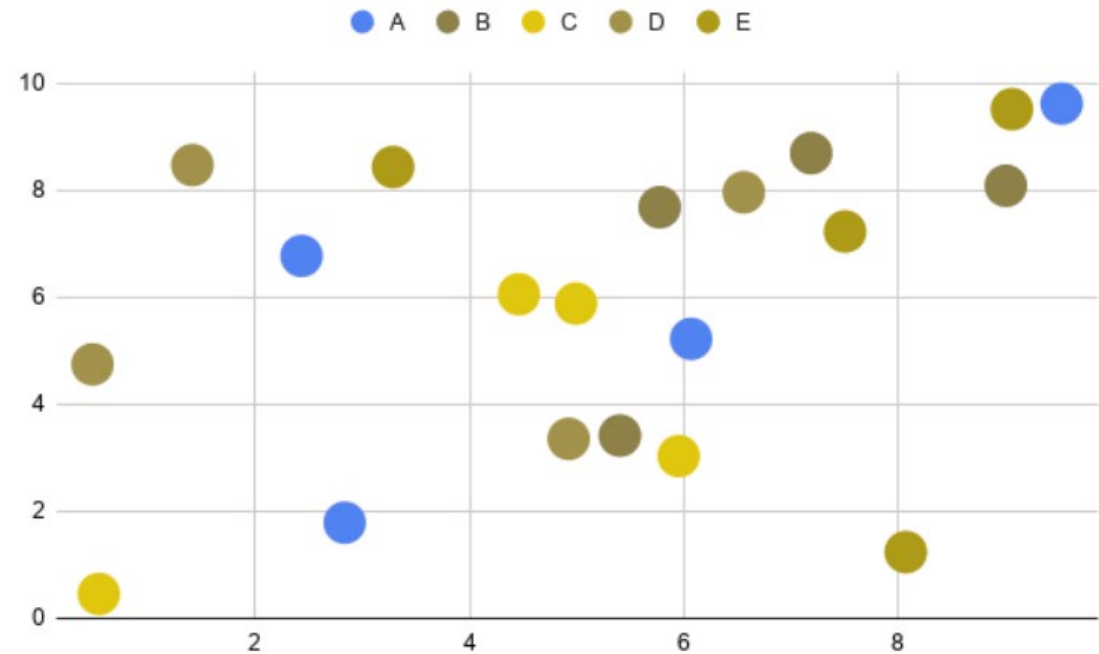
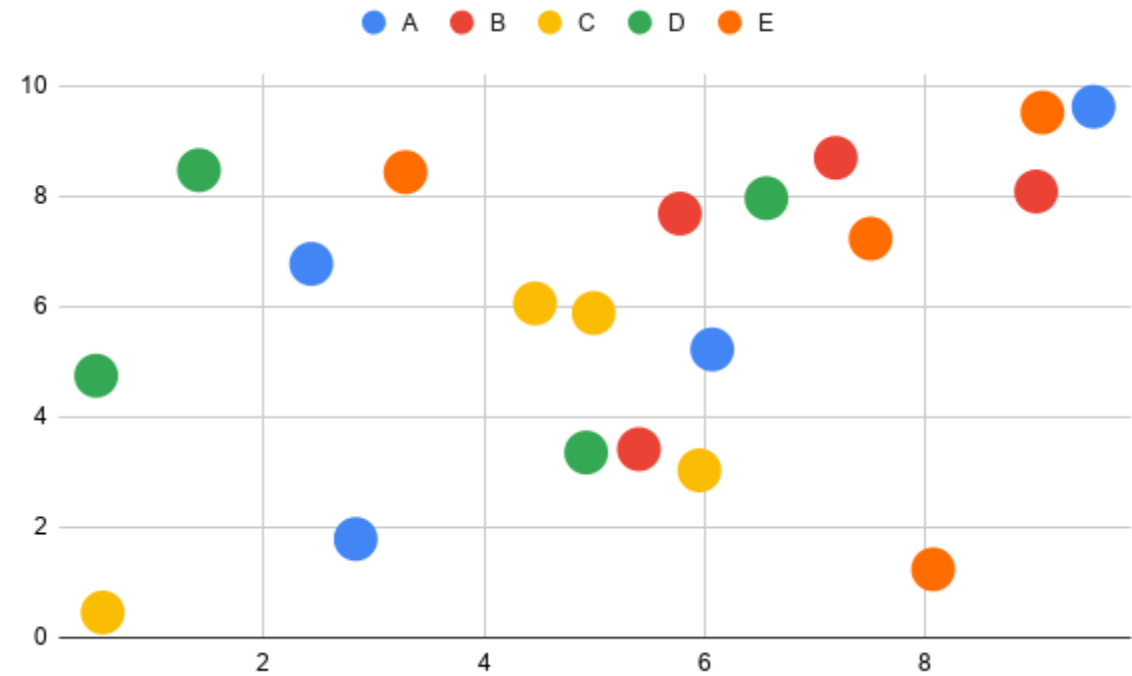
COLOR BLIND TEST

Do you see a number inside each circle?



Colour-blind

- Selective
- Associative
- Quantitative
- Order
- Variations / Length / Resolution



Colour-blind adjusted scales

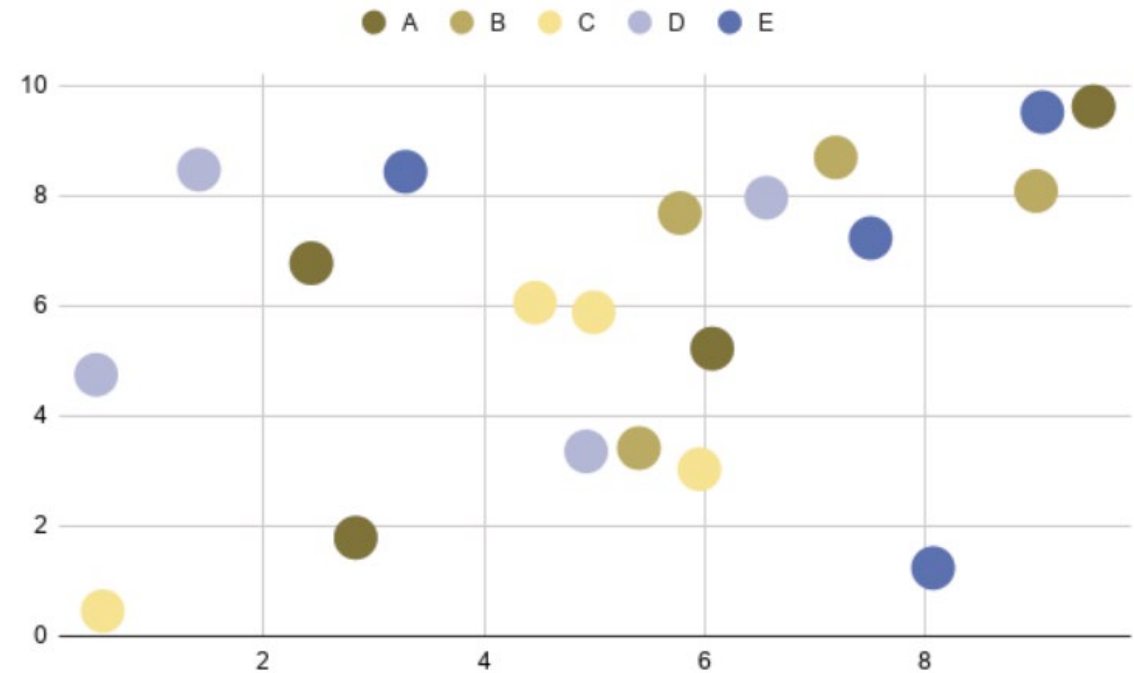
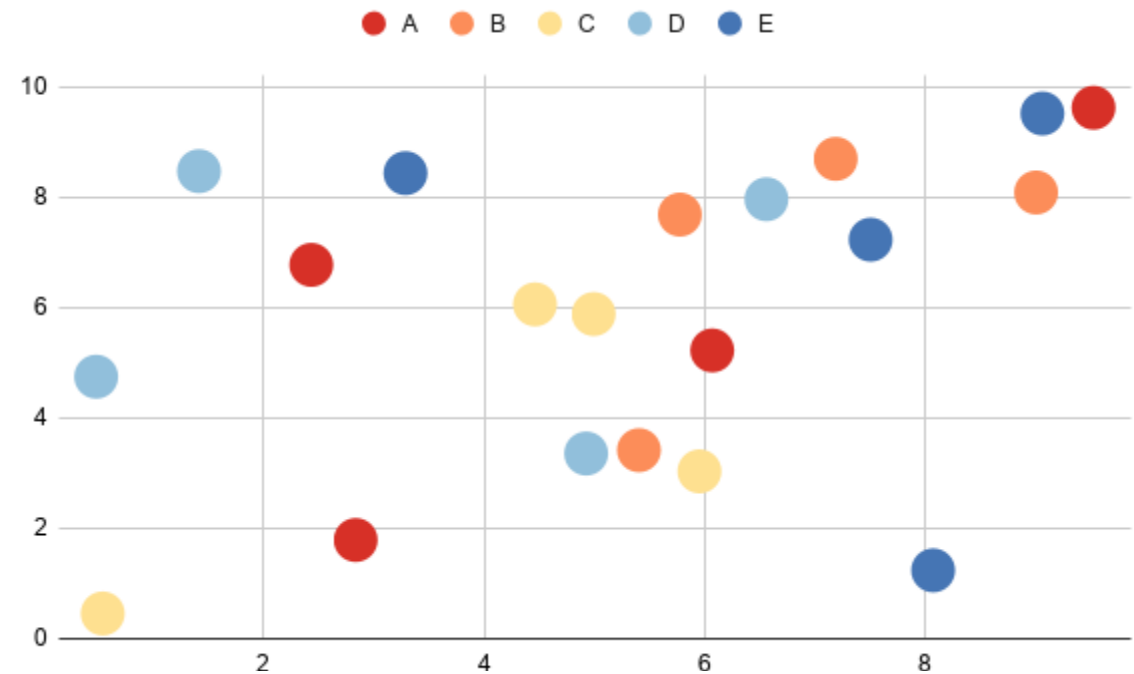
+/- Selective

+/- Associative

- Quantitative

- Order

+/- Variations / Length / Resolution



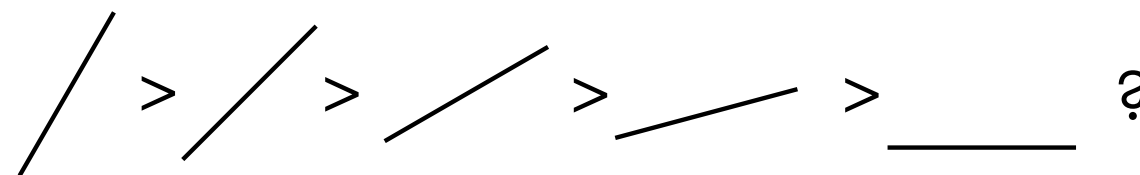
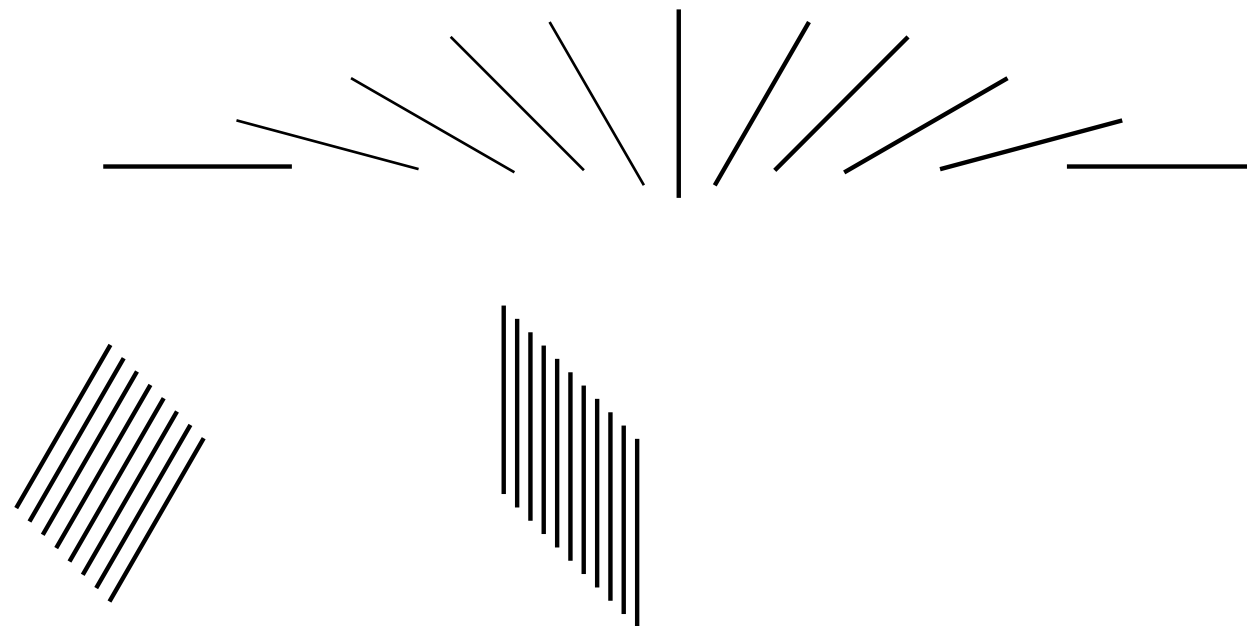
Orientation

Selective
Associative

Quantitative

Order

Variations / Length / Resolution



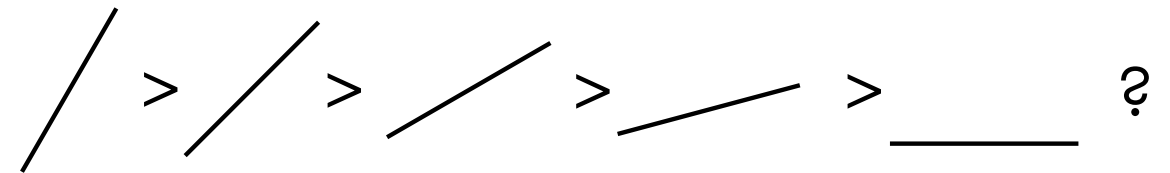
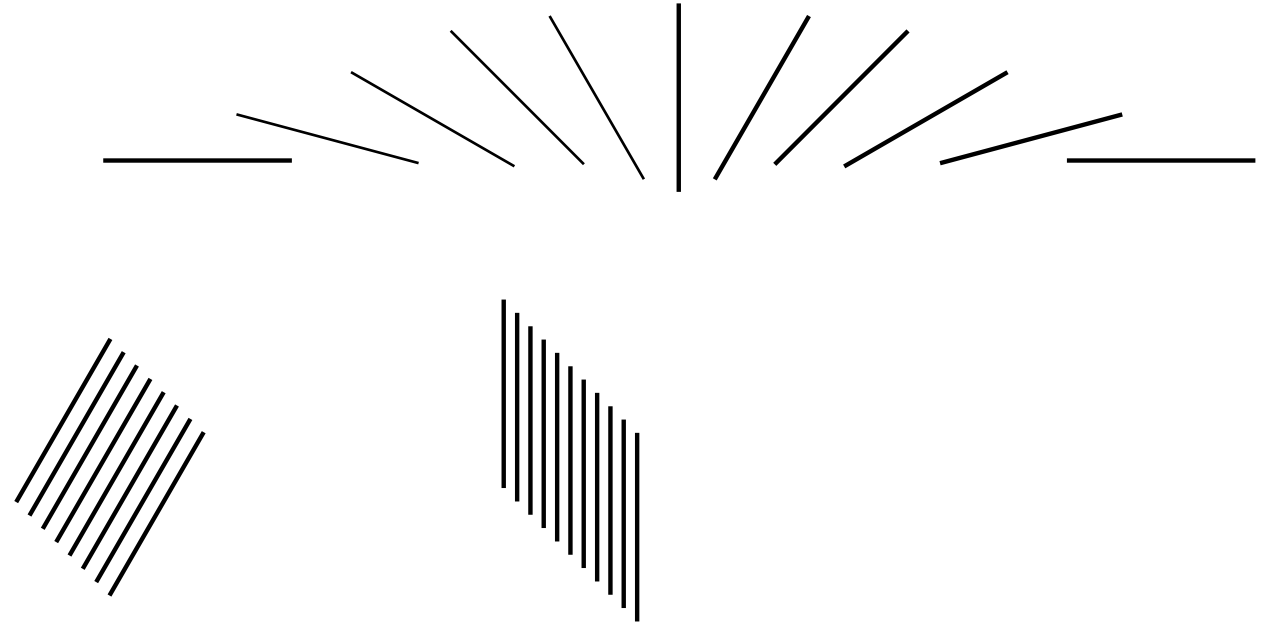
Orientation

+ Selective
+ Associative

- Quantitative

- Order

+ Variations / Length / Resolution (~5 in 2D, ? In 3D)



Texture

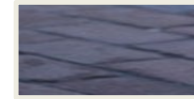
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Texture

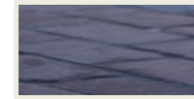
+ Selective

+ Associative

- Quantitative

- Order

+ Variations / Length / Resolution (theoretically infinite)



Carpendale 2003

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

Visual Hierarchy

1. Reading patterns (many left->right, scan patterns F and Z)
2. Size dictates focus order
3. Space (texture) – emphasis
4. Type-Face – bold emphasis, italics supplemental
5. Colour – colour important, b/w distance
6. Direction – grids common structure, but breaking grid can pull focus

Onward to ... Obtaining Data

Jonathan Hudson
jwhudson@ucalgary.ca
<https://pages.cpsc.ucalgary.ca/~jwhudson/>



UNIVERSITY OF
CALGARY