Visualization

DATA 201: Thinking With Data
Winter 2021

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

Monday, January 25, 2021

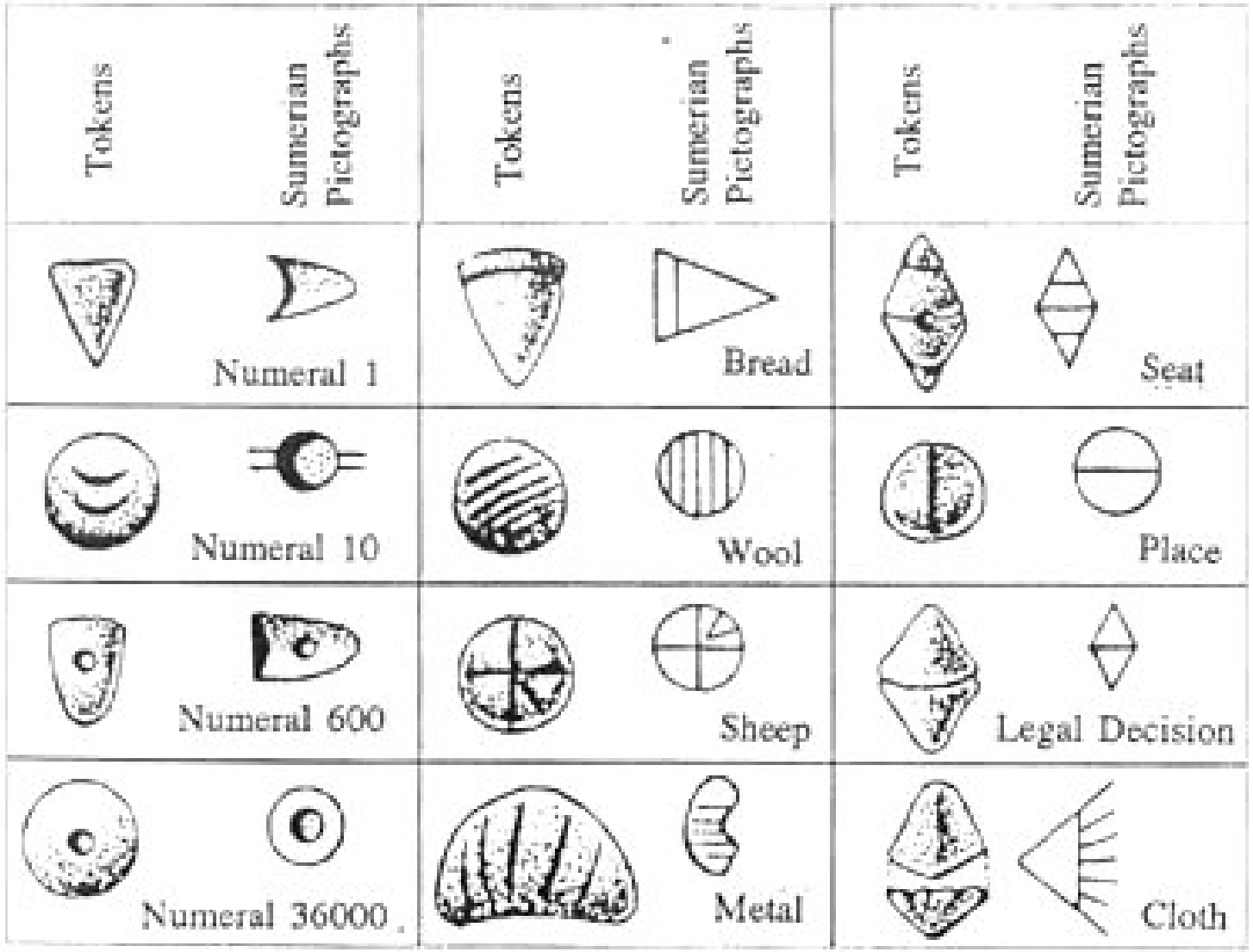
UNIVERSITY OF CALGARY

What is visualization?

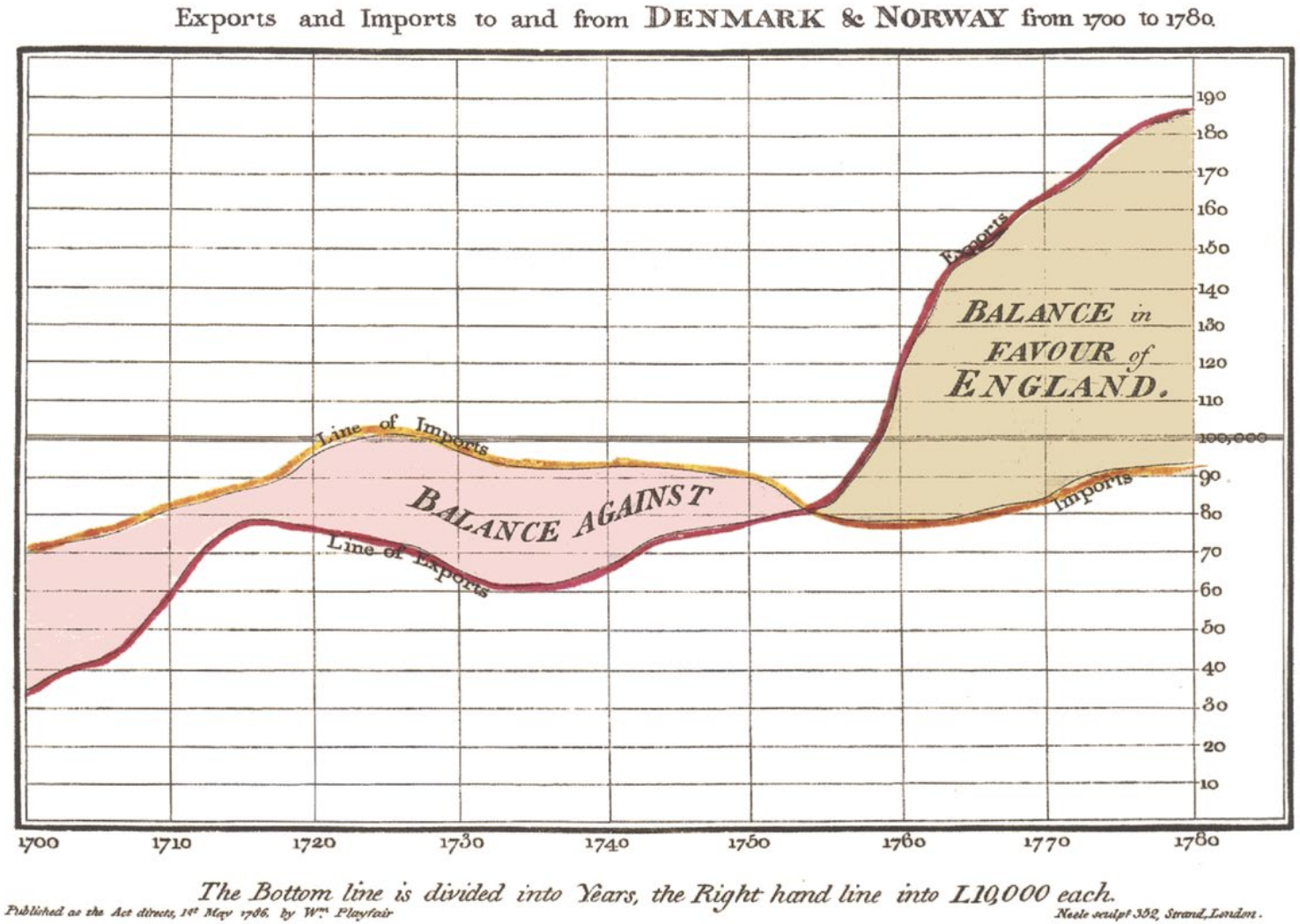
Scientific Visualization
Information Visualization
Data Visualization

Mesopotamian Clay Tokens 5500 BCE

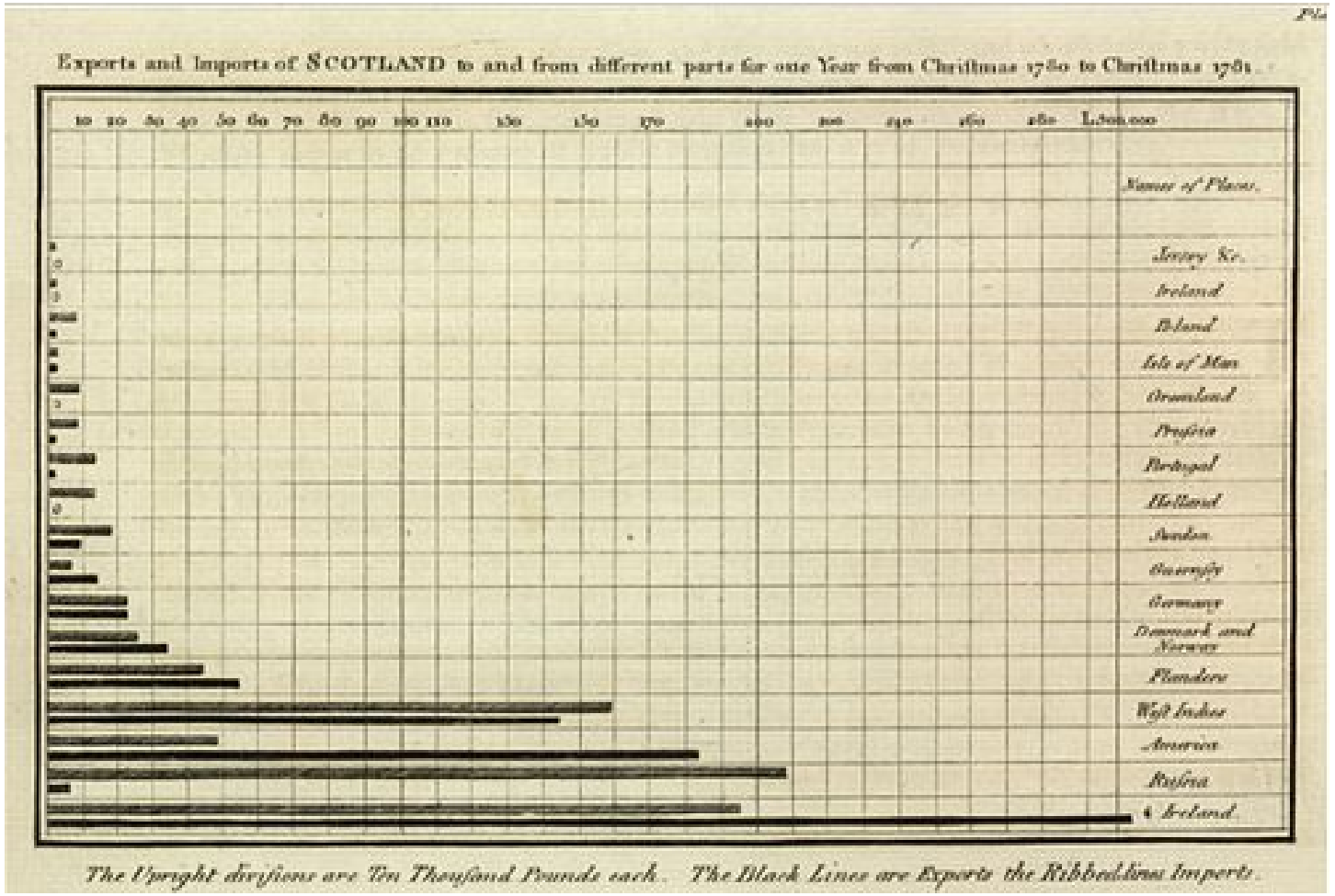
Size -> Quantity
Shape -> Type

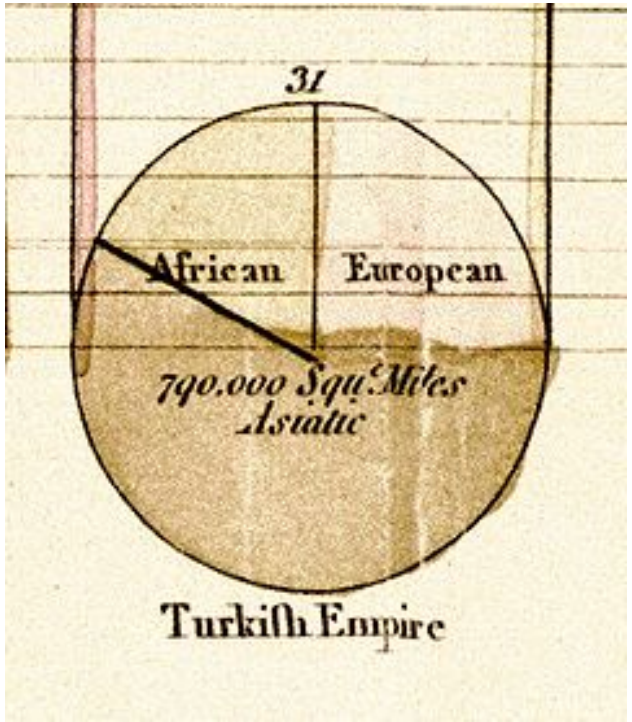


William Playfair 1786



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





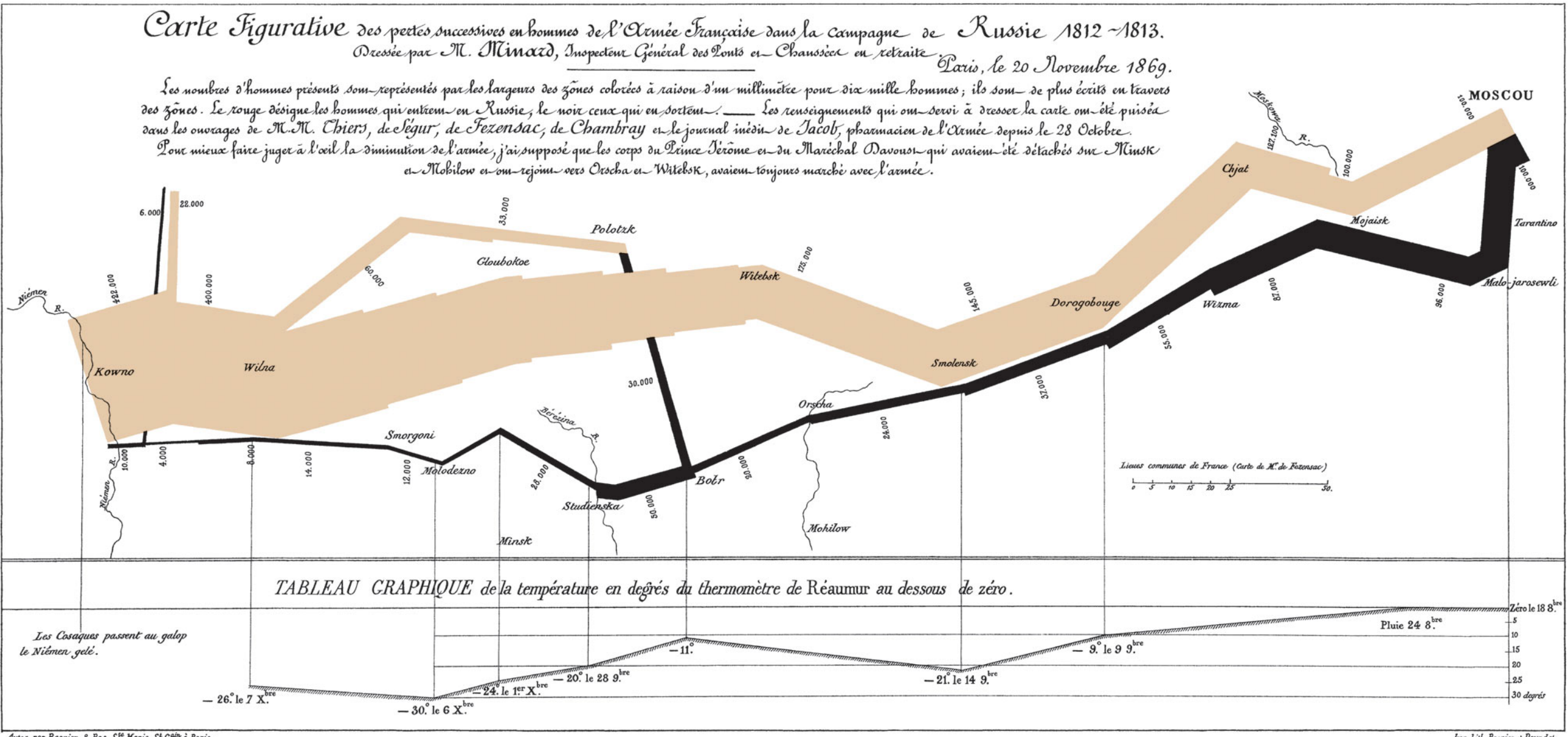
John Snow 1854

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



Charles Joseph Minard 1869

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

Charles Joseph Minard 1869

COVID

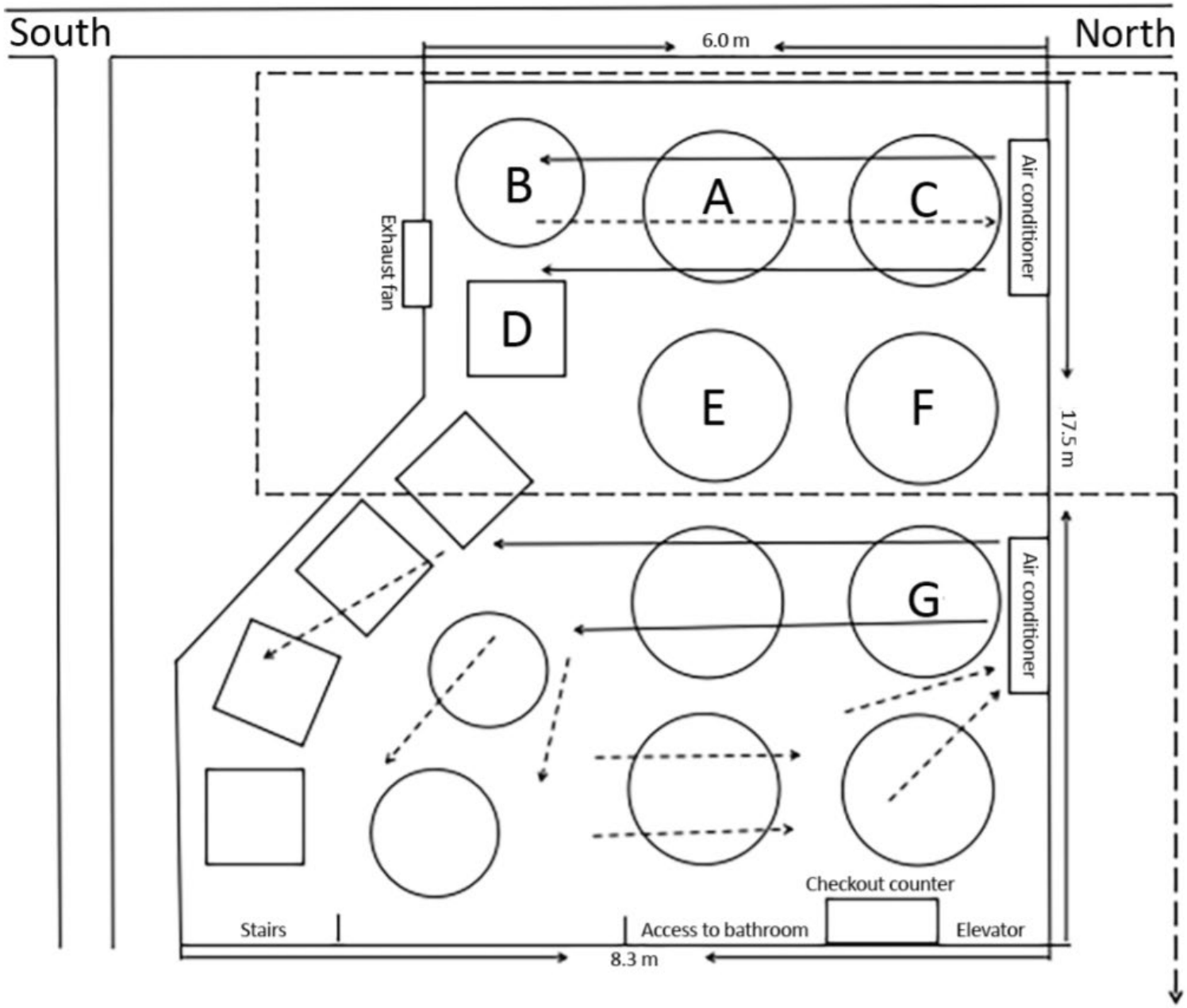


Figure. Sketch showing arrangement of restaurant tables and air conditioning airflow at site of outbreak of 2019 novel coronavirus disease, Guangzhou, China, 2020. Red circles indicate seating of future case-patients; yellow-filled red circle indicates index case-patient.

- CDC https://wwwnc.cdc.gov/eid/article/26/7/20-0764_article

COVID

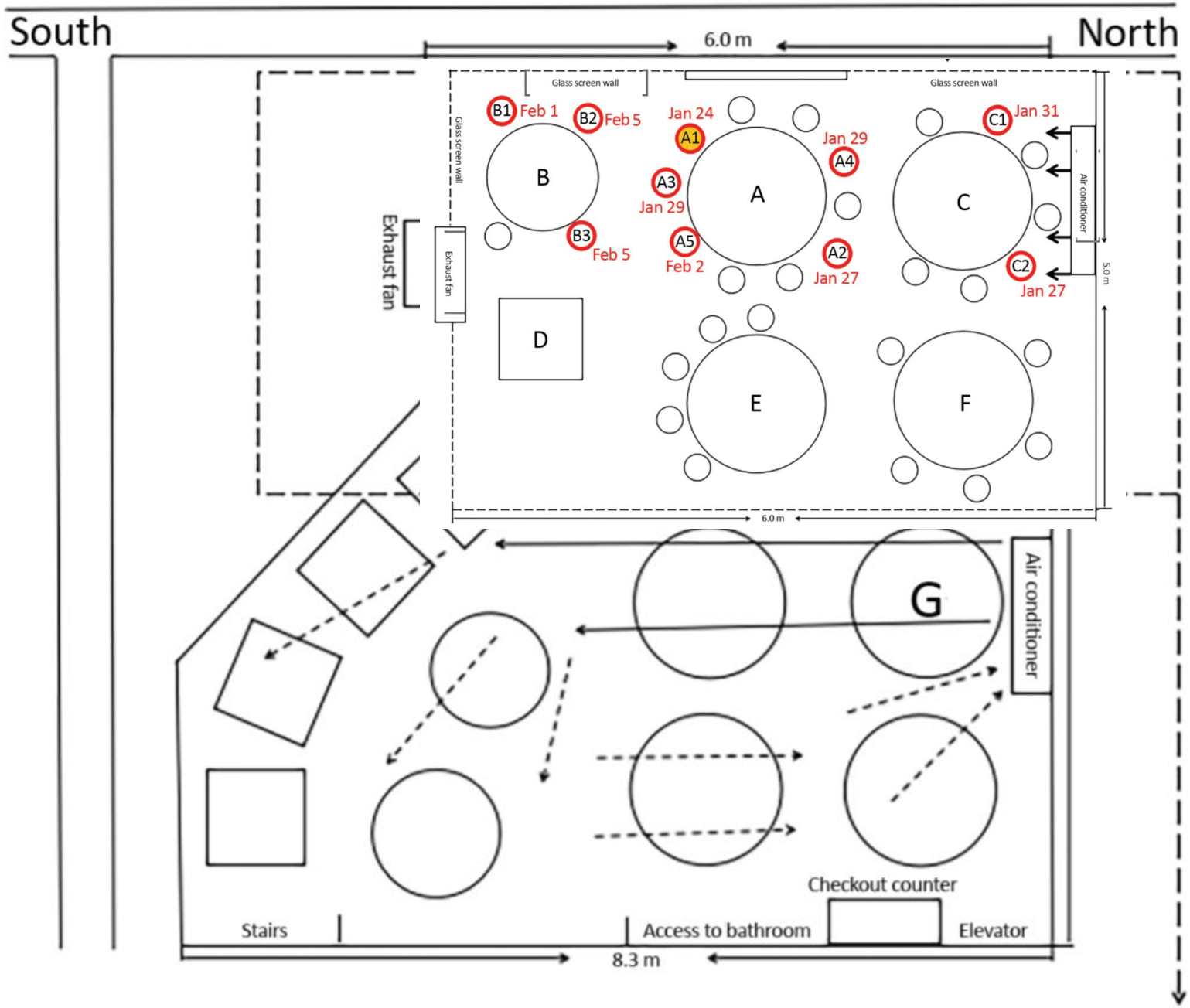
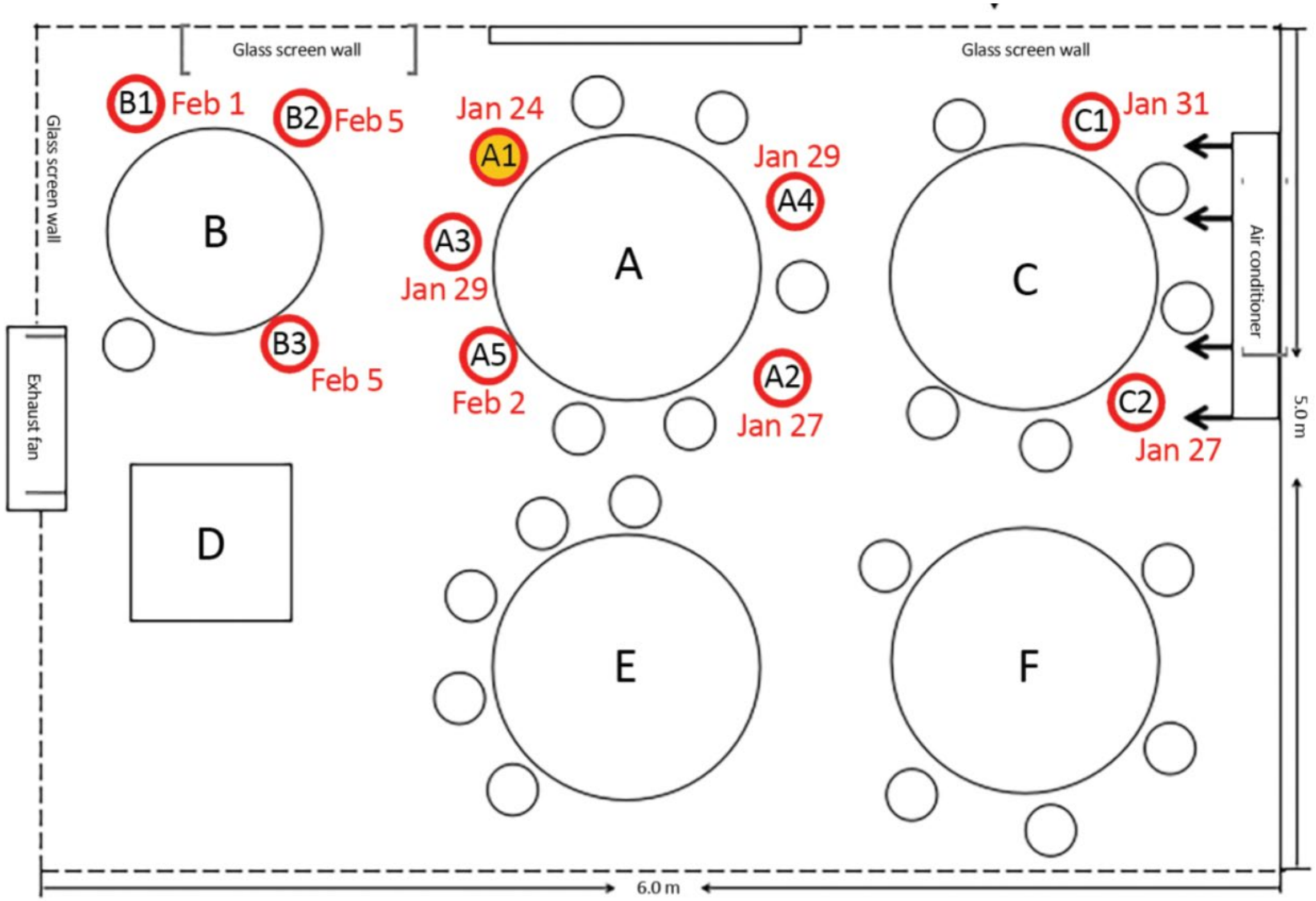


Figure. Sketch showing arrangement of restaurant tables and air conditioning airflow at site of outbreak of 2019 novel coronavirus disease, Guangzhou, China, 2020. Red circles indicate seating of future case-patients; yellow-filled red circle indicates index case-patient.

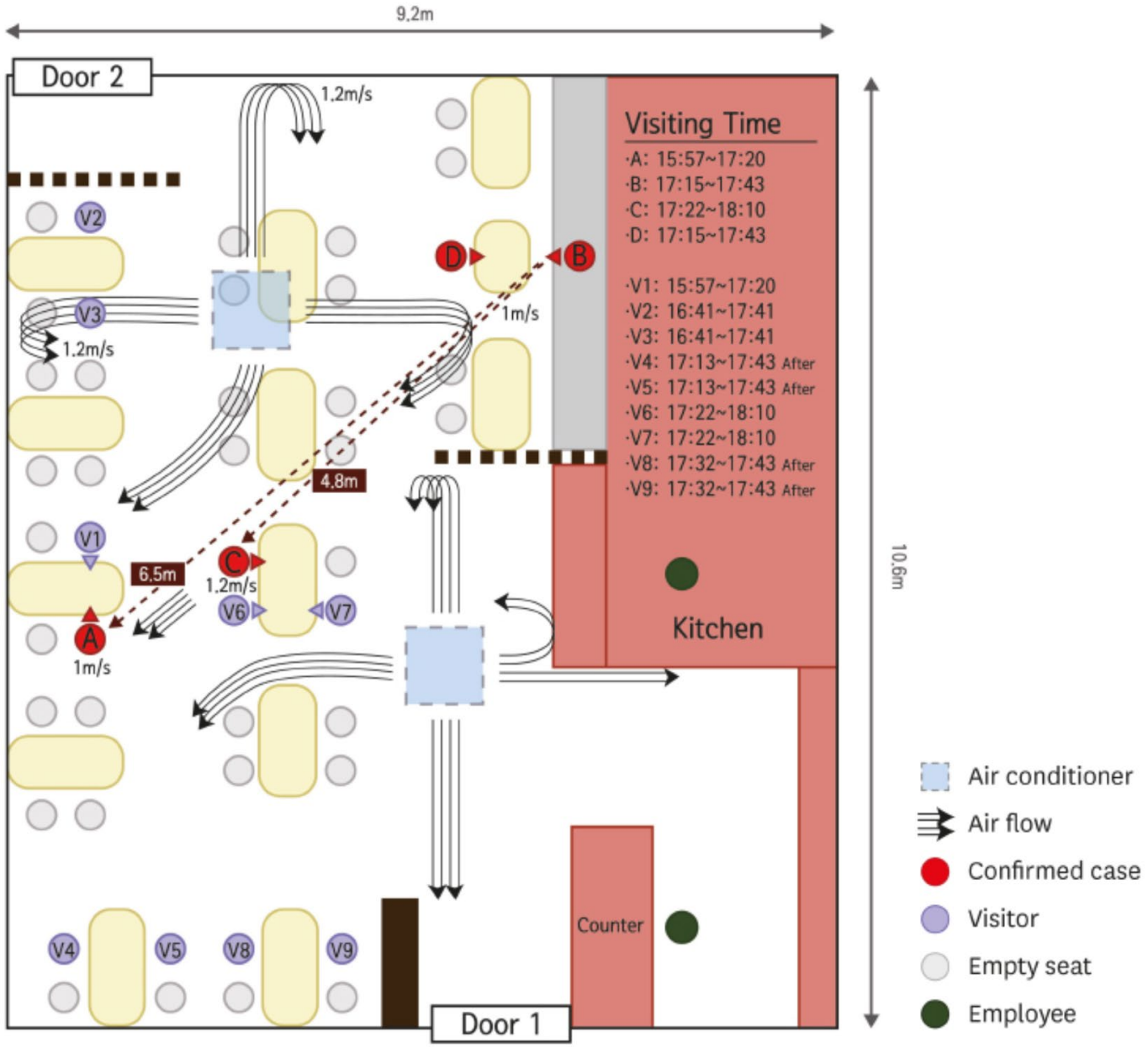
- CDC https://wwwnc.cdc.gov/eid/article/26/7/20-0764_article

COVID



- April 19th, 2020

6 Feet Not Enough



- JKMS https://jkms.org/DOIx.php?id=10.3346/jkms.2020.35.e415

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 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 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 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 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 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 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 me 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

Lines

Areas

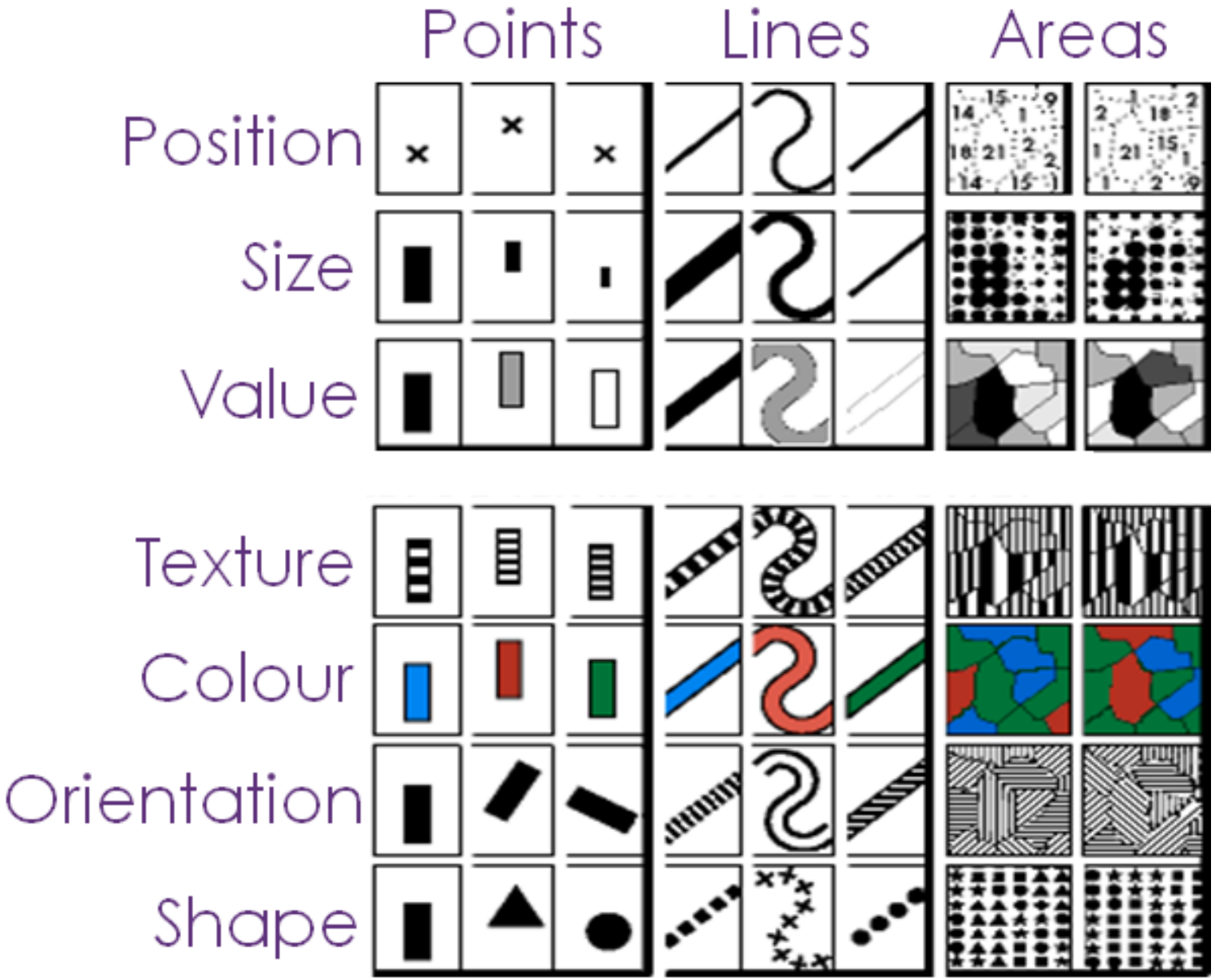
Bertin 1974

Visual Variables

Visual Variables

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

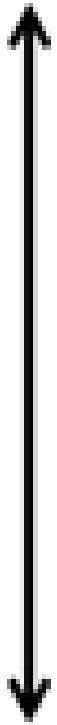
Bertin 1974



UNIVERSITY OF CALGARY

	Nominal	Ordinal	Quantitative
Position			
Size			
Value			
Texture			
Colour			
Orientation			
Shape			

More Accurate



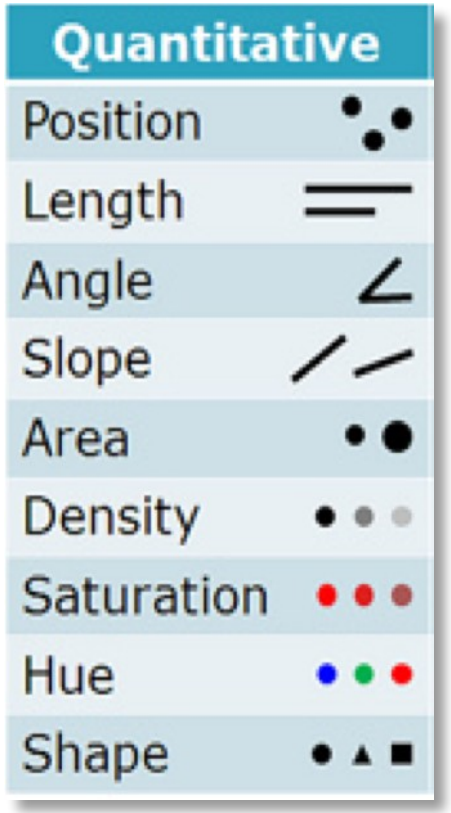
Less Accurate

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

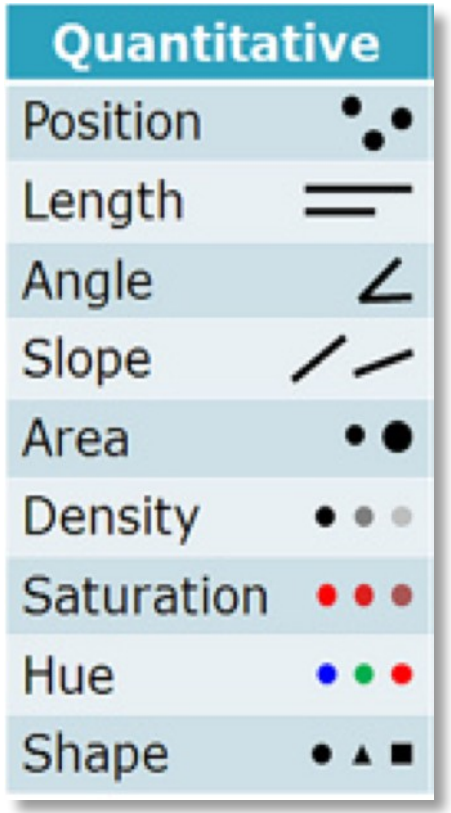
Jacques Bertin refined by Cleveland & McGill then by Card & Mackinlay

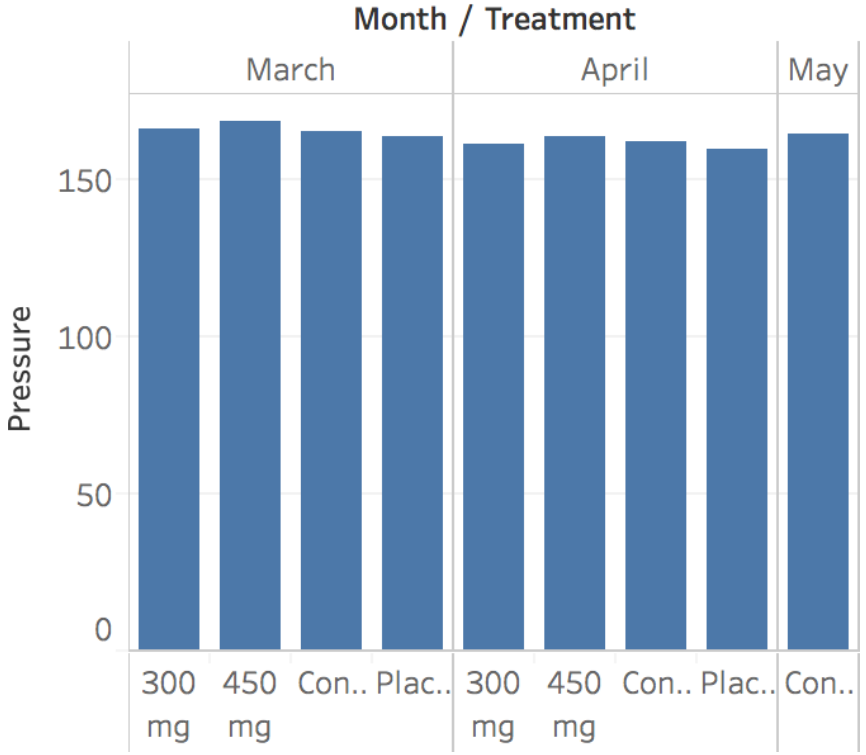
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

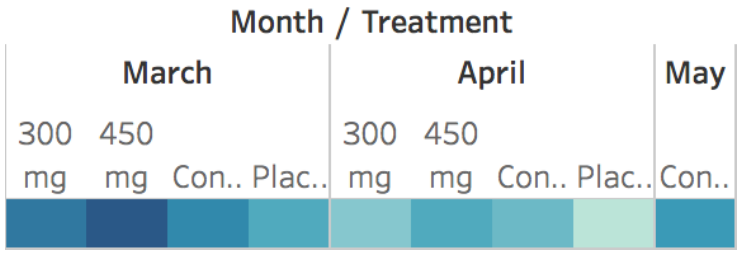
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

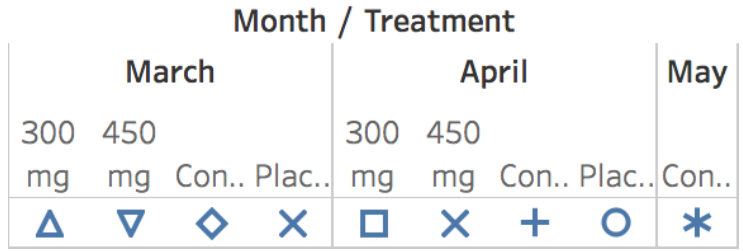


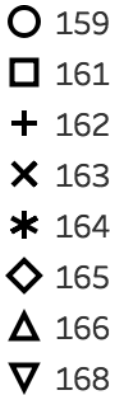
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164





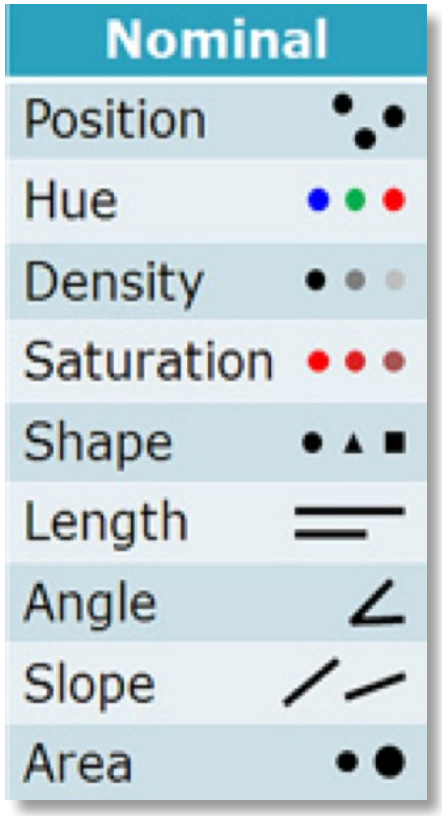




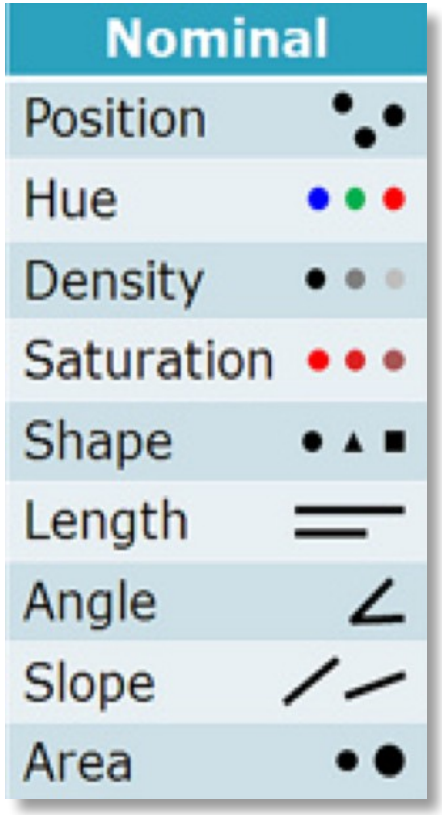


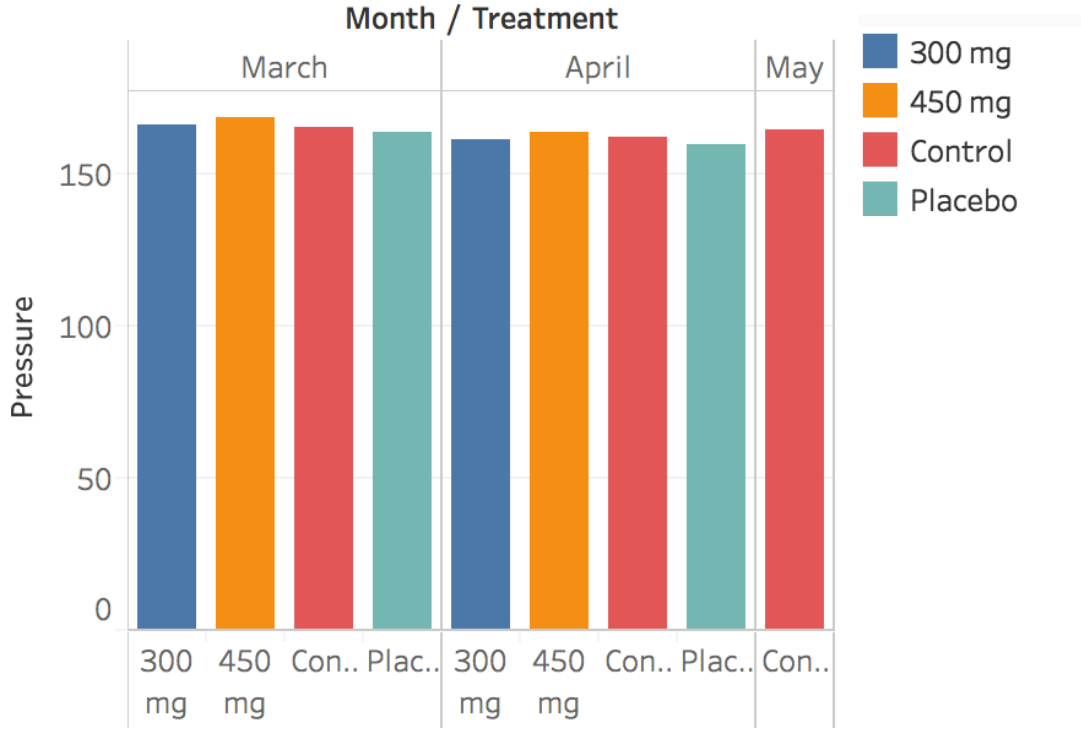
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

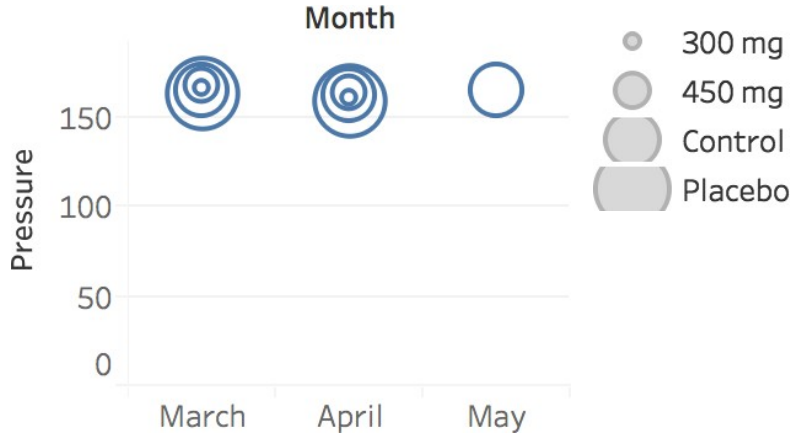
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164



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







UNIVERSITY OF CALGARY

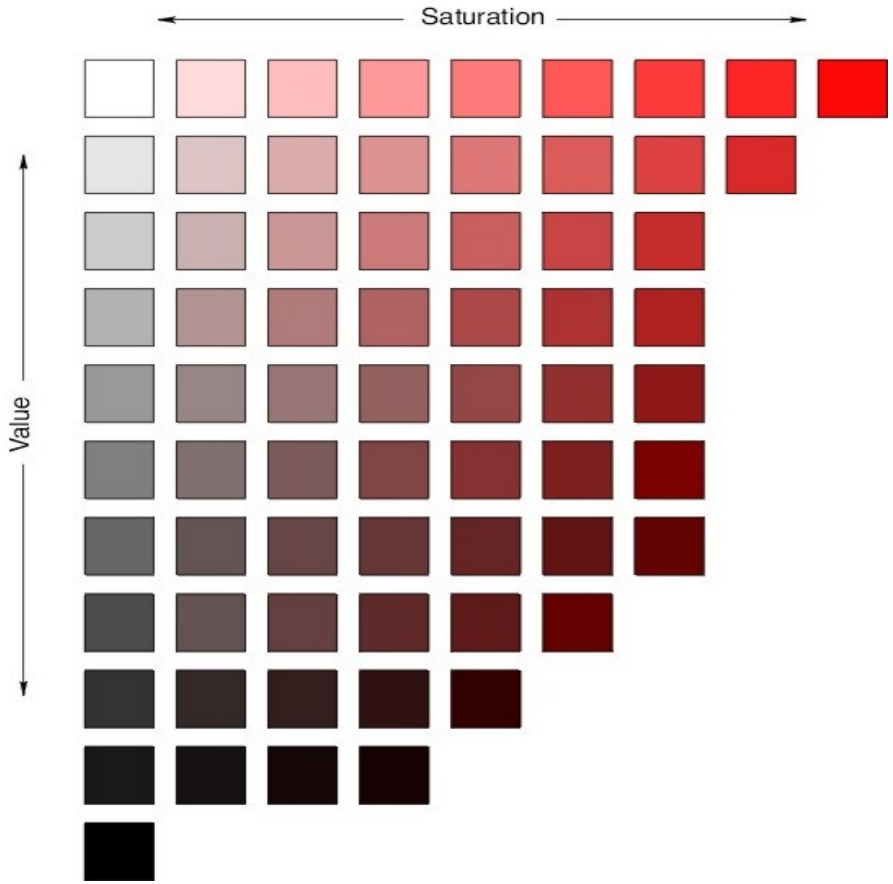
More Variables with Computers

Motion

More Variables with Computers

Motion

Colour
- Hue (actual colour)
- Value (brightness)
- Saturation (intensity)



More Variables with Computers

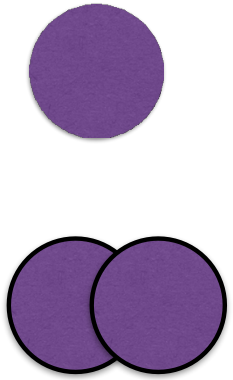
Flicker
  Frequency, rhythm

More Variables with Computers

Flicker

Frequency, rhythm

Depth

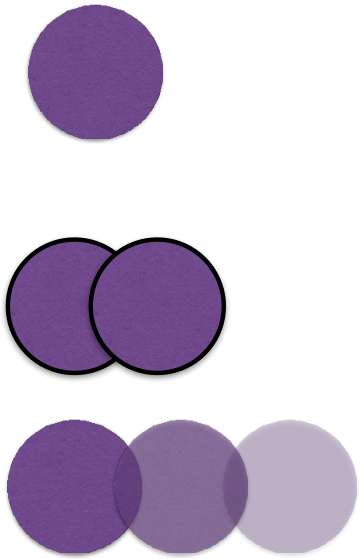


More Variables with Computers

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

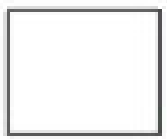
- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order
- Variations / Length / Resolution

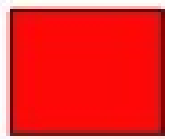
Characteristics of Visual Variables

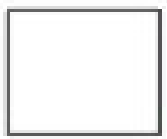
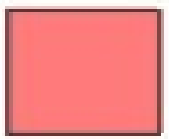
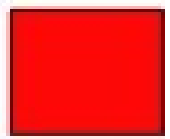
- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order — perceive an order
- Variations / Length / Resolution

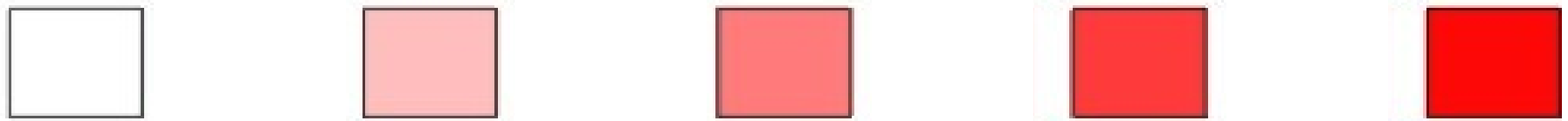
Characteristics of Visual Variables

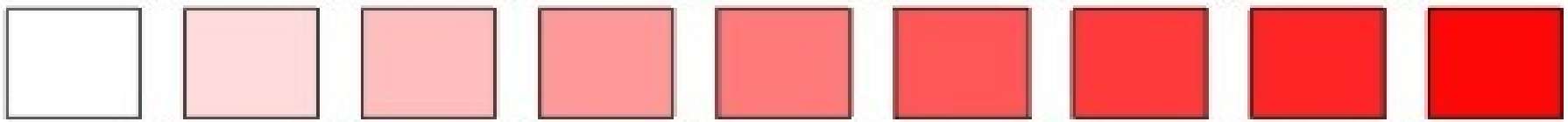
- Selective — differentiate items from groups
- Associative — group items in a group
- Quantitative — changes in terms of numerical reading
- Order — perceive an order
- Variations / Length / Resolution — distinguishable variations (How many variations in the visual variable are distinctions recognizable?)









Position

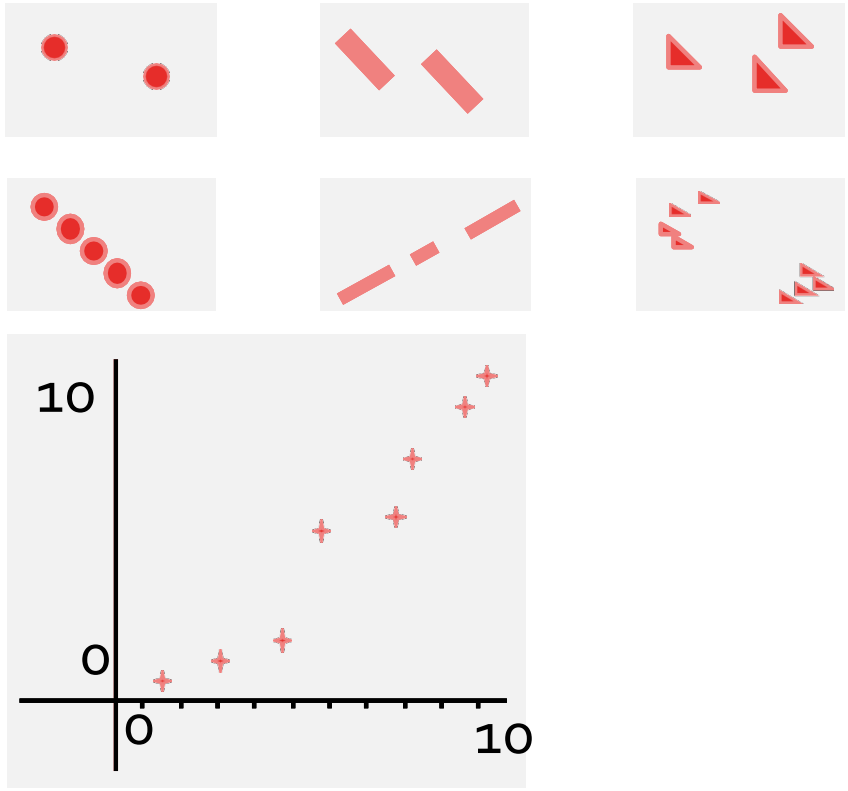
Selective

Associative

Quantitative

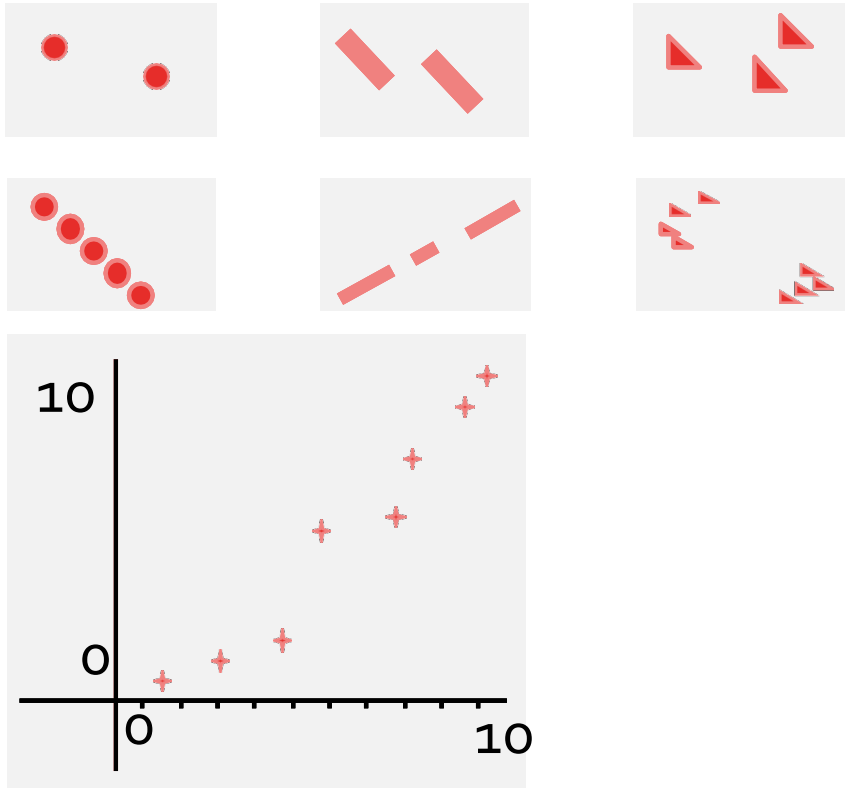
Order

Variations / Length / Resolution



Position

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



Size

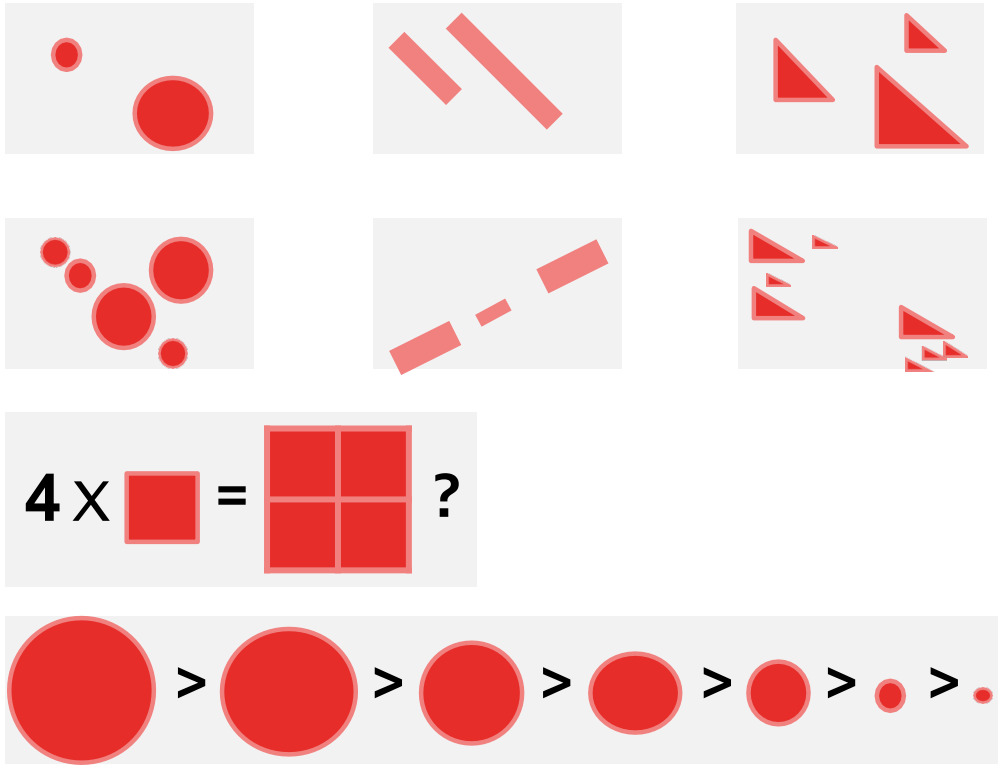
Selective

Associative

Quantitative

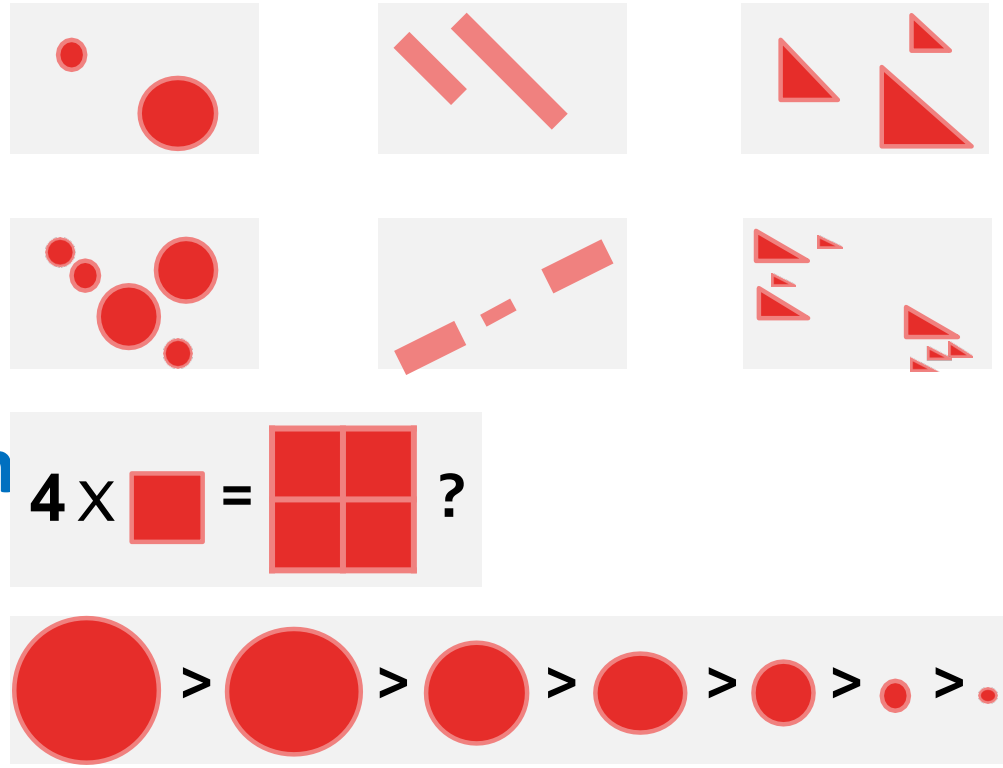
Order

Variations / Length / Resolution



Size

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



Shape

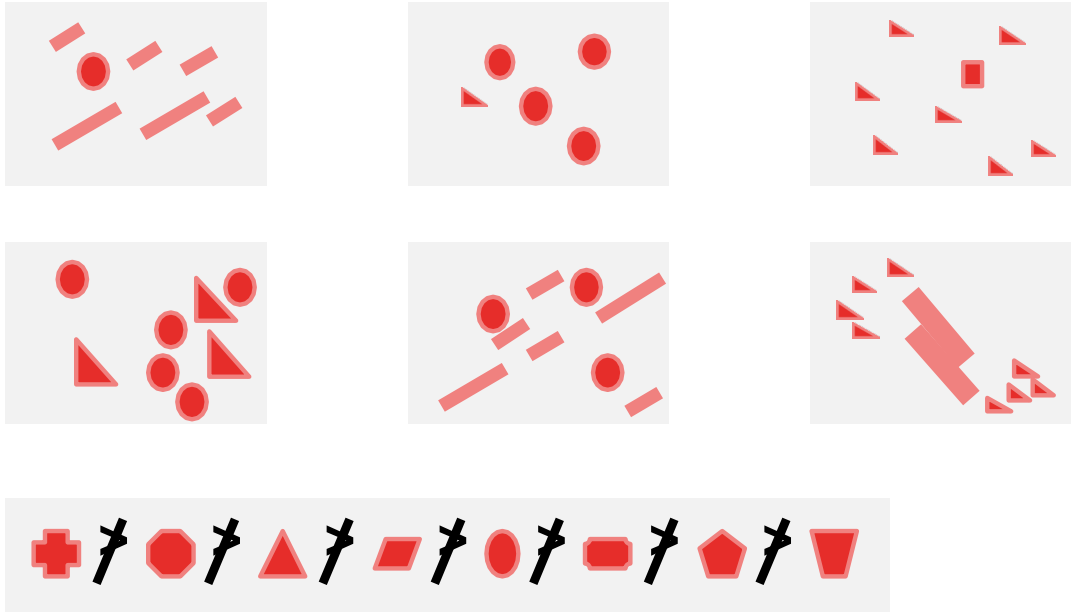
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Shape

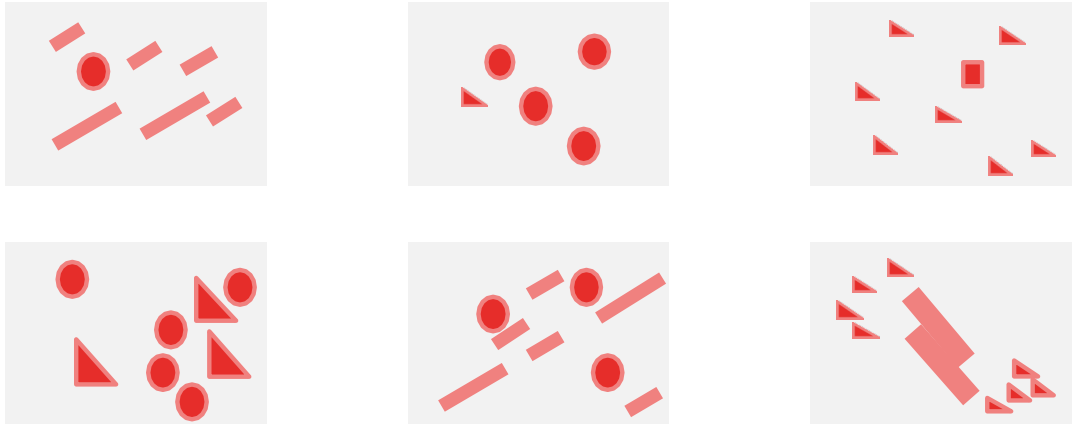
+/- Selective

+/- Associative

- Quantitative

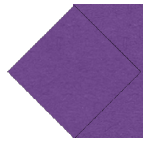
- Order

+ Variations / Length / Resolution

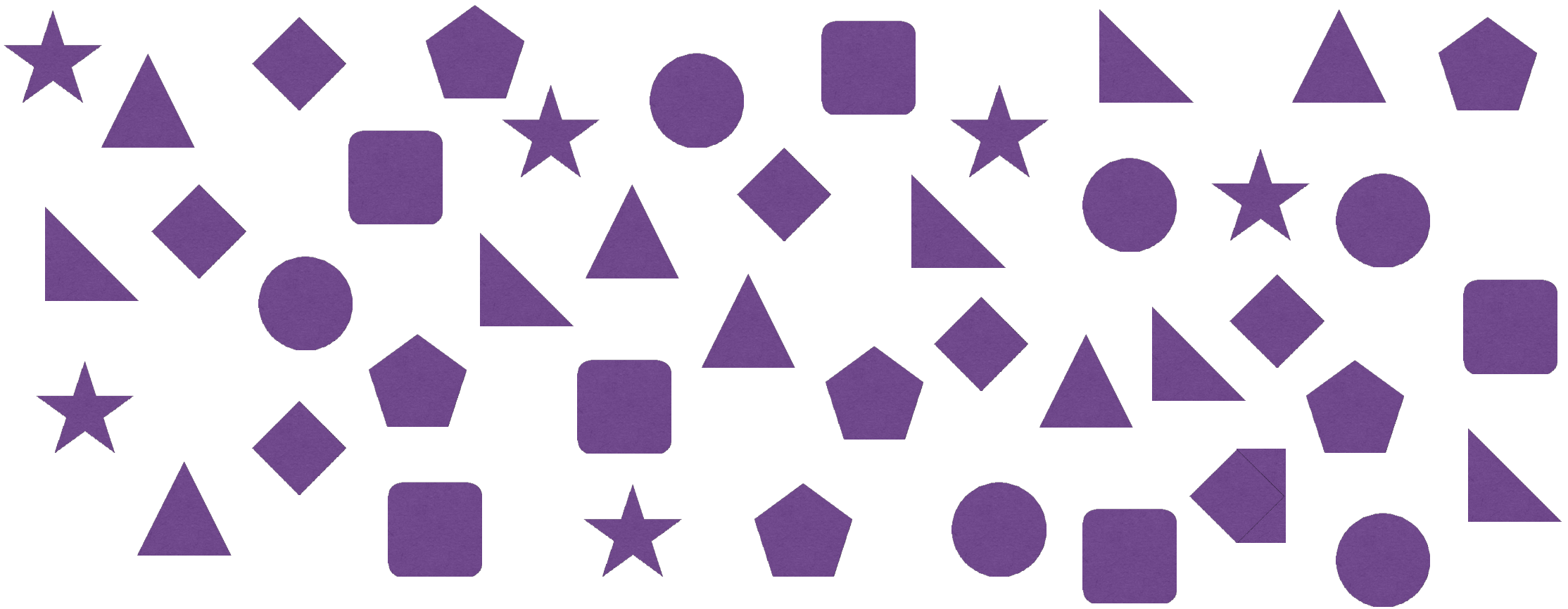


Shape

Find



Shape



Value

Selective

Associative

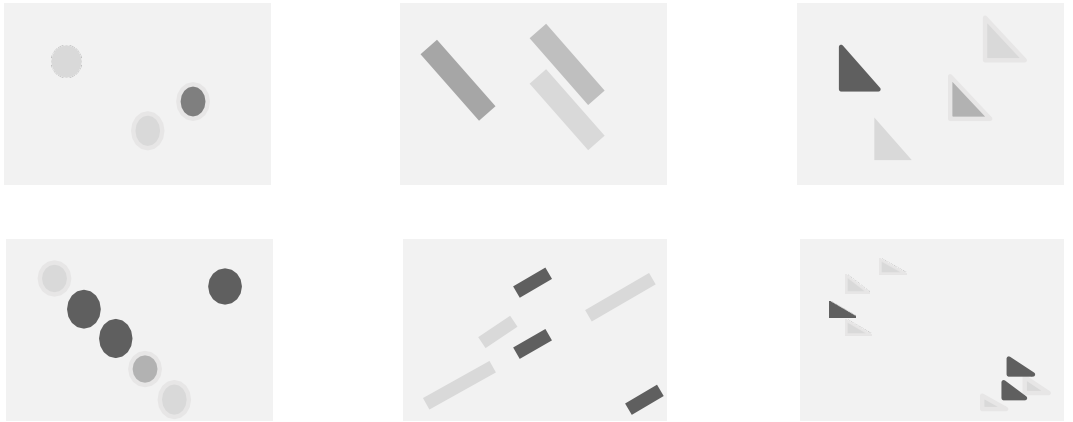
Quantitative

Order

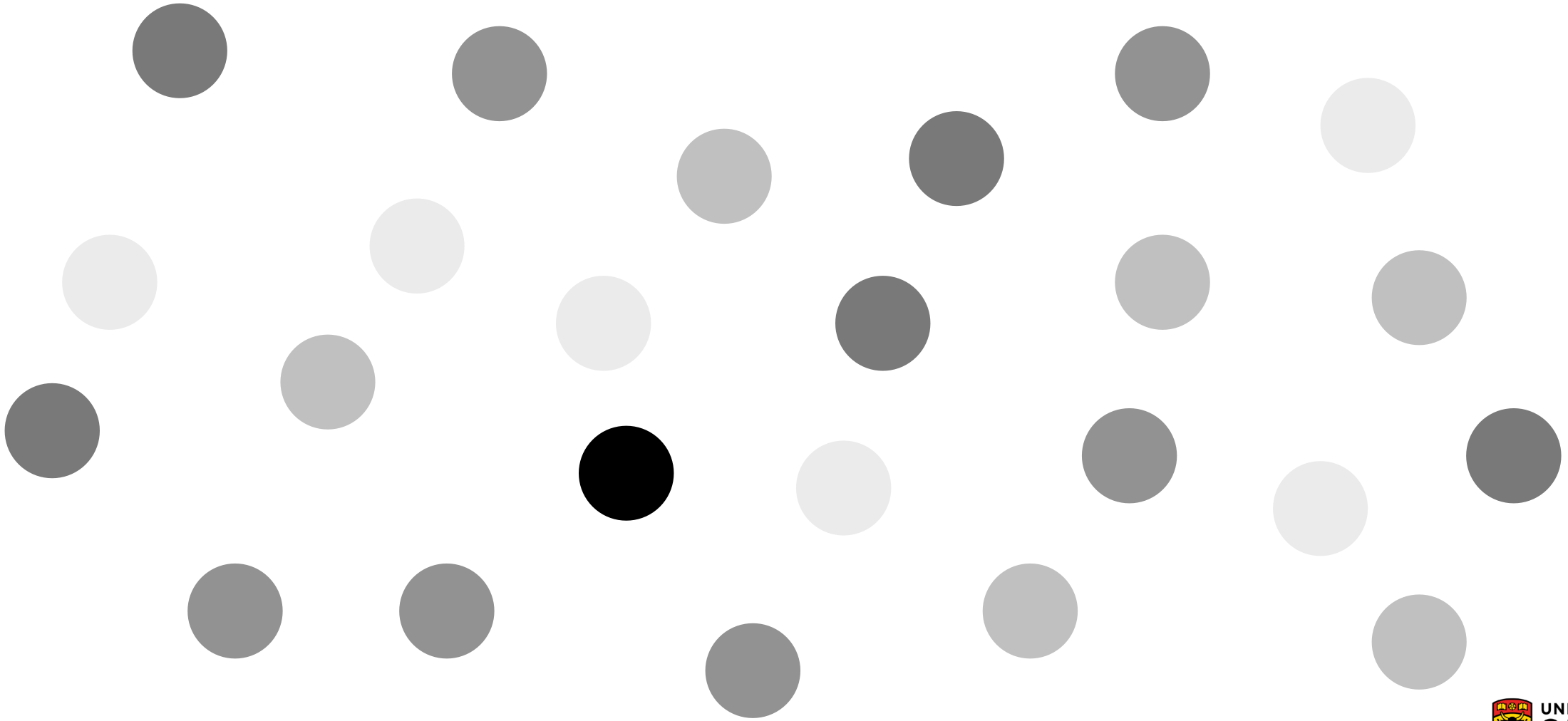
Variations / Length / Resolution

Value

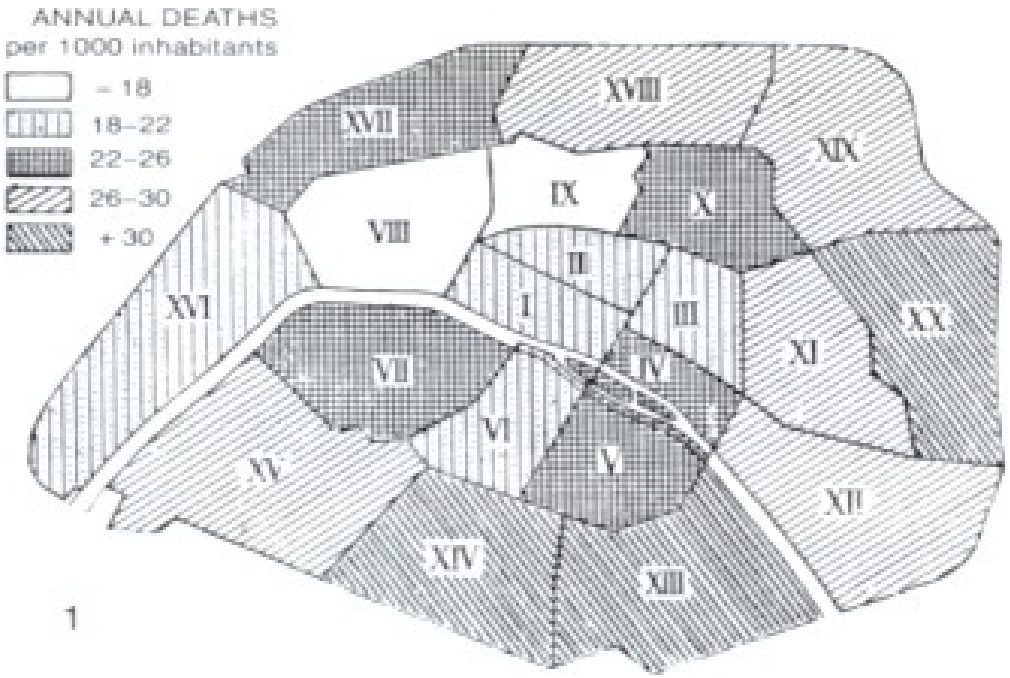
+ Selective
+ Associative
+/- Quantitative
+ Order
+ Variations / Length / Resolutic

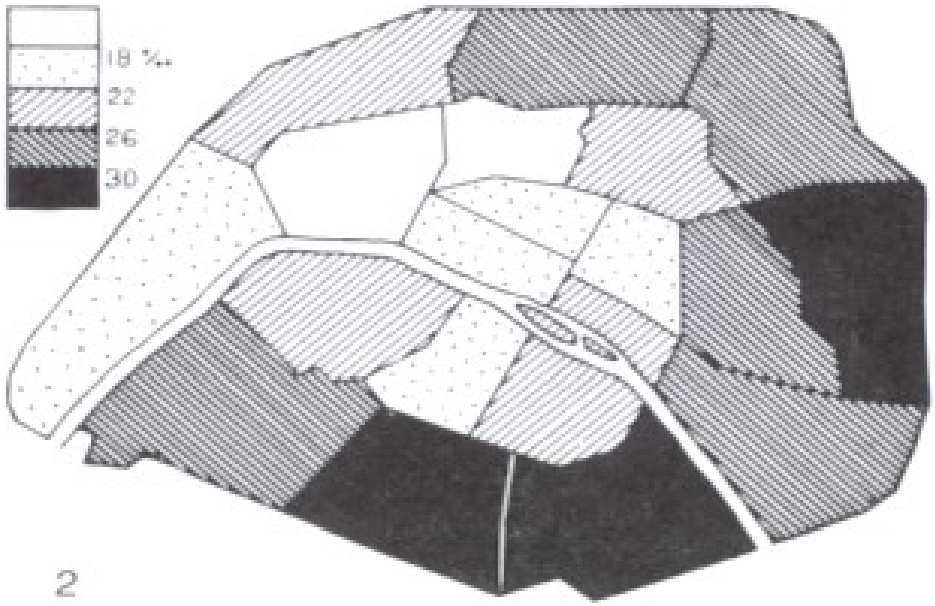


Value



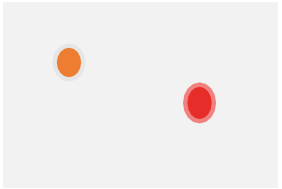
Value



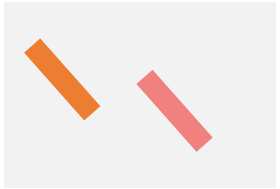


Colour

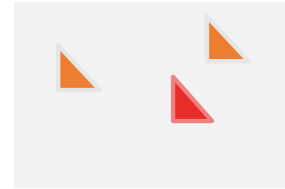
Selective



Associative

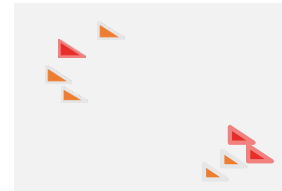


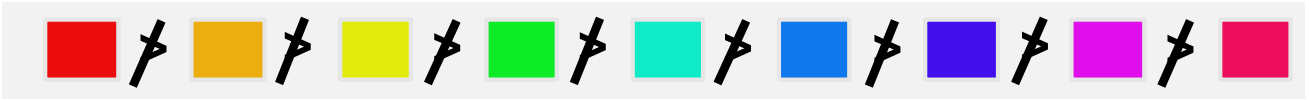
Quantitative



Order

Variations / Length / Resolution





Colour

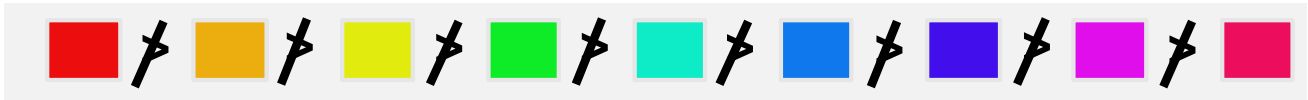
+ Selective

+ Associative

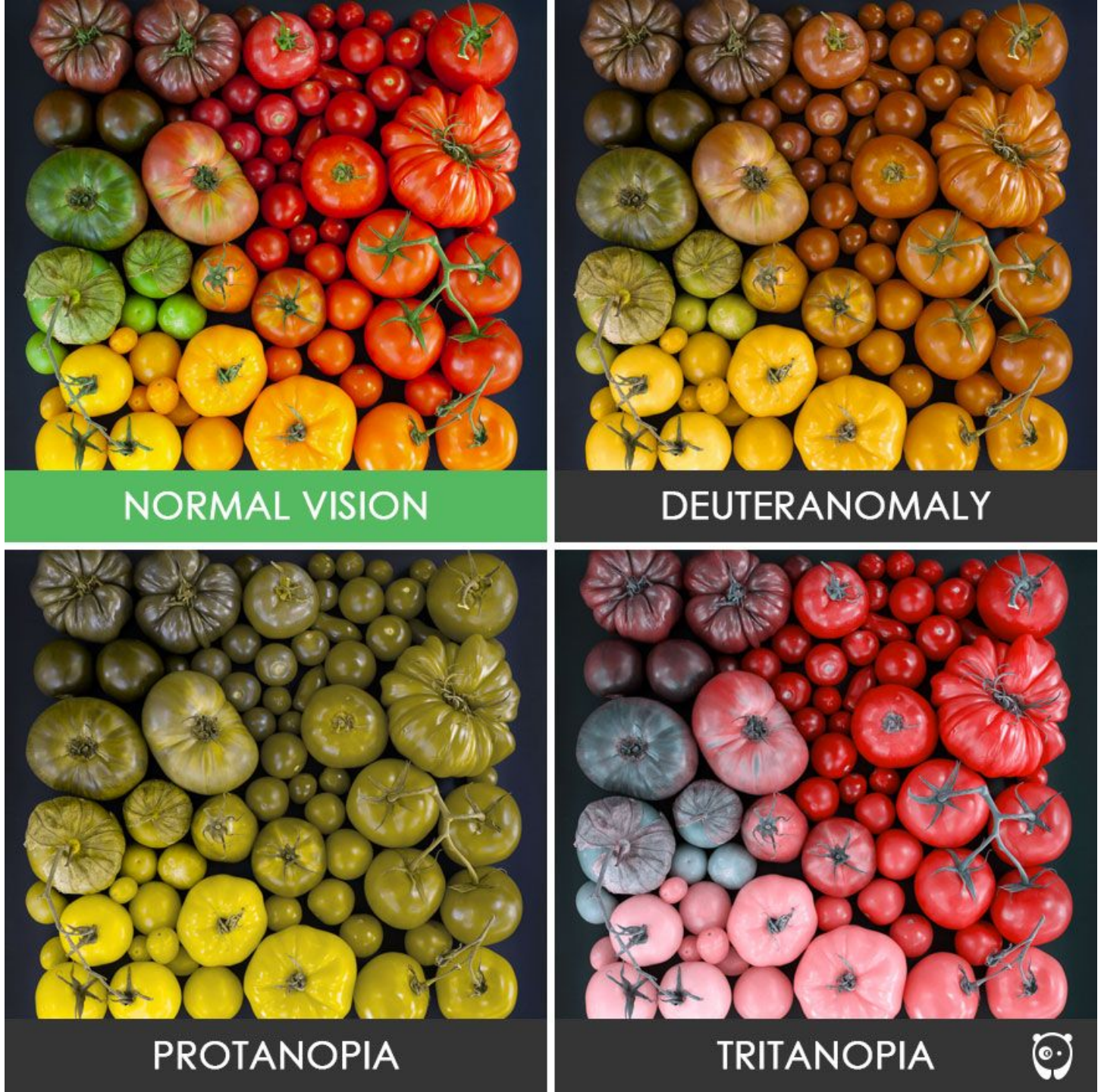
- Quantitative

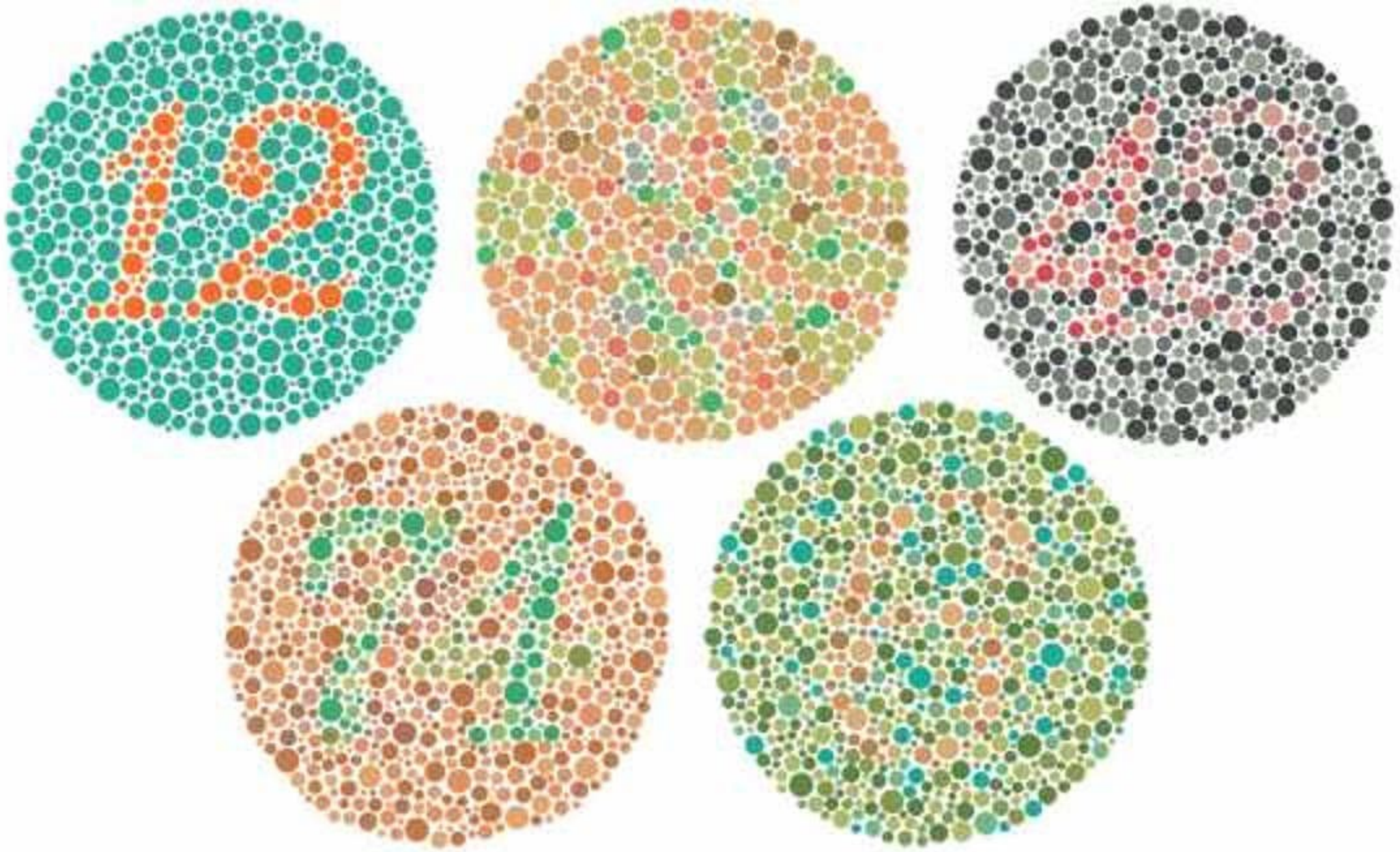
- Order

+ Variations / Length / Resolution



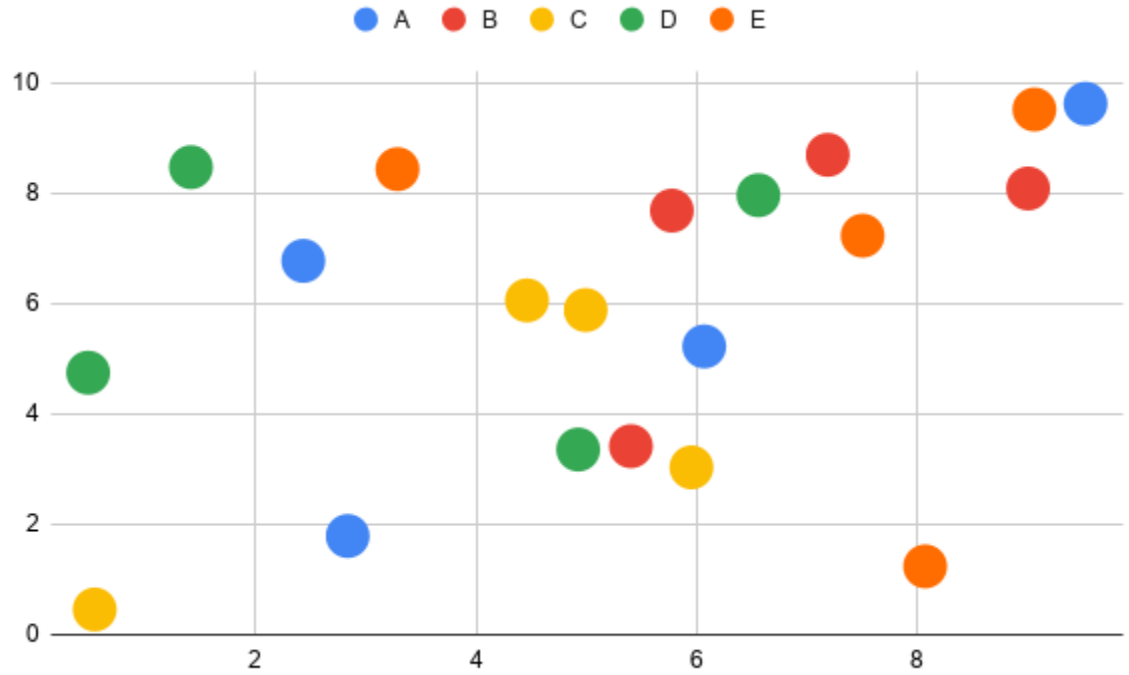
Colour-blind

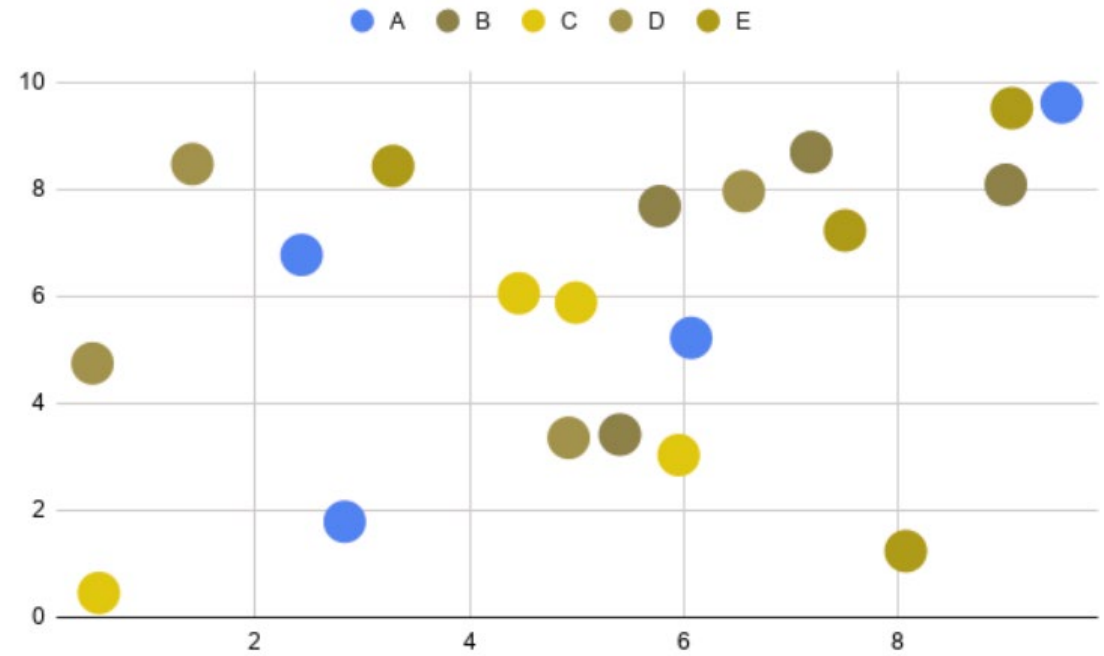




Colour-blind

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





Colour-blind

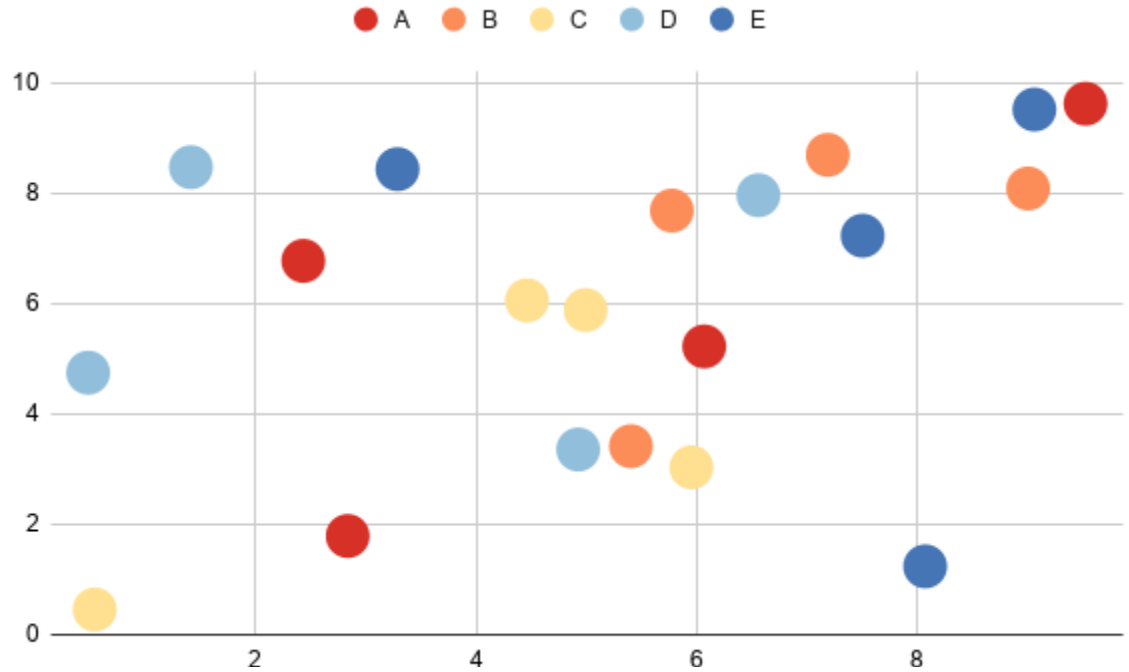
+/- Selective

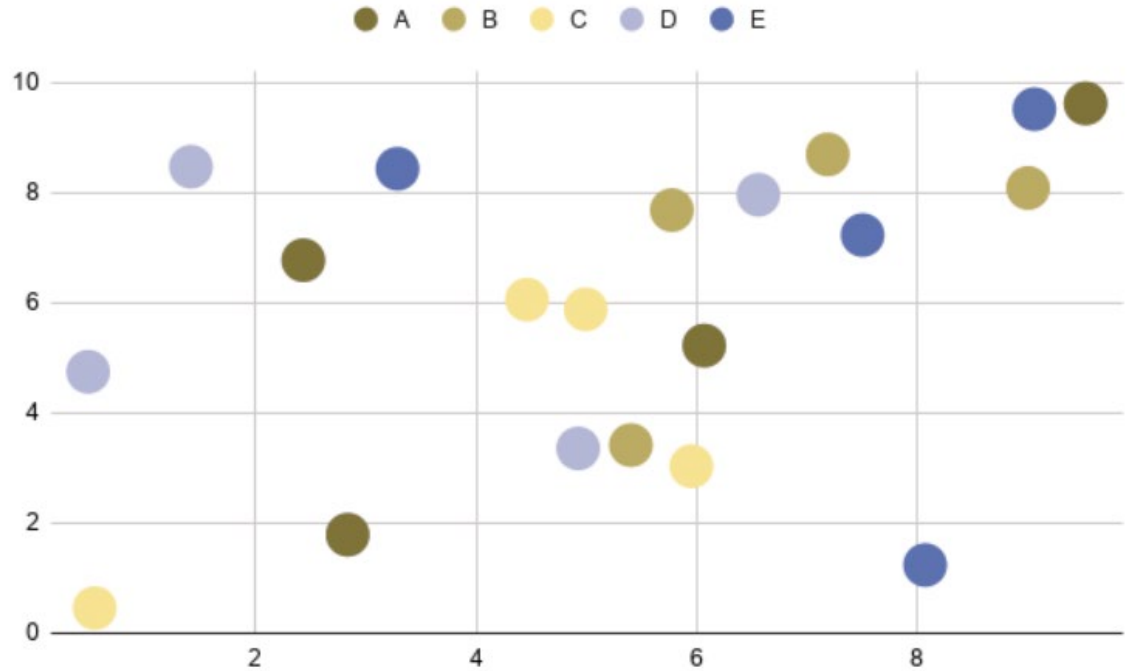
+/- Associative

- Quantitative

- Order

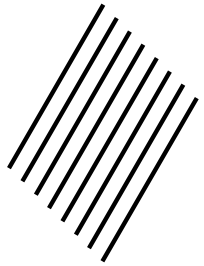
+/- Variations / Length / Resolution





Orientation

Selective

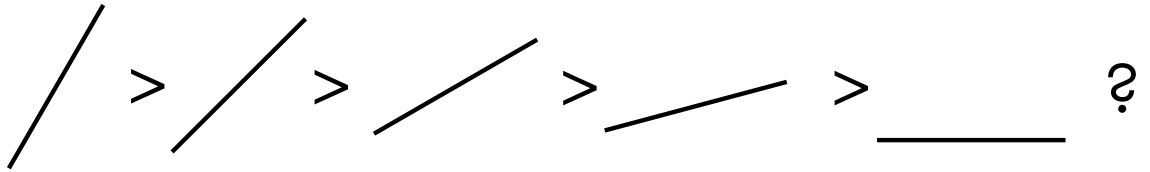


Associative

Quantitative

Order

Variations / Length / Resolution



Orientation

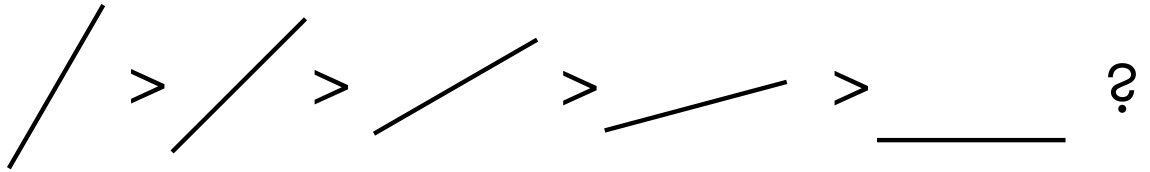
+ Selective

+ Associative

- Quantitative

- Order

+ Variations / Length / Resolution



Texture

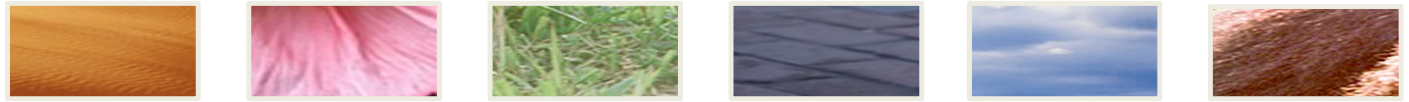
Selective

Associative

Quantitative

Order

Variations / Length / Resolution



Texture

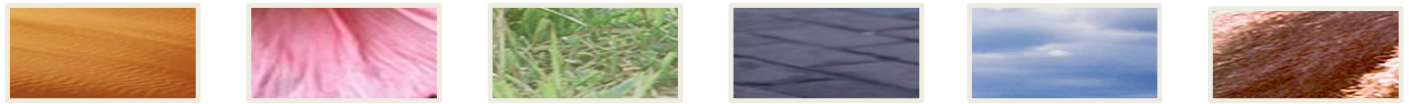
+ Selective

+ Associative

- Quantitative

- Order

+ Variations / Length / Resolution



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