

TRỰC QUAN HÓA DỮ LIỆU

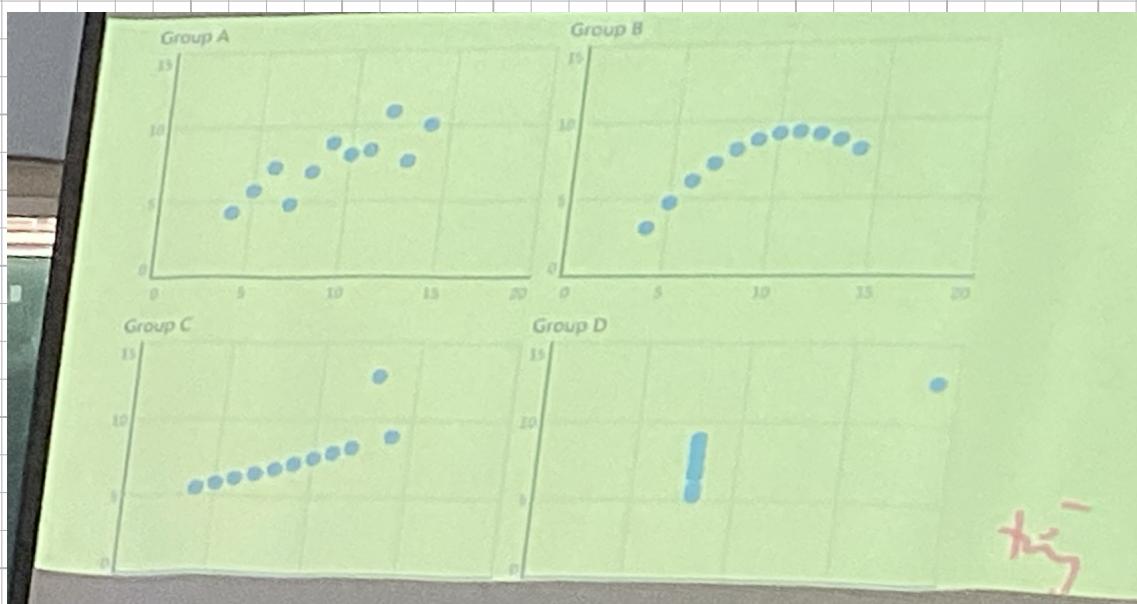
60% {
 15% - Seminar (giữa kỳ)
 30% - văn phòng (cuối kỳ)
 15% - group (hoạt động trên Zalo)}

40% : Thực hành LAB

Trực quan là quá trình trình bày chuyển đổi liệu thô thành

đồ liệu hình ảnh để dễ dàng tiếp nhận nhanh chóng.

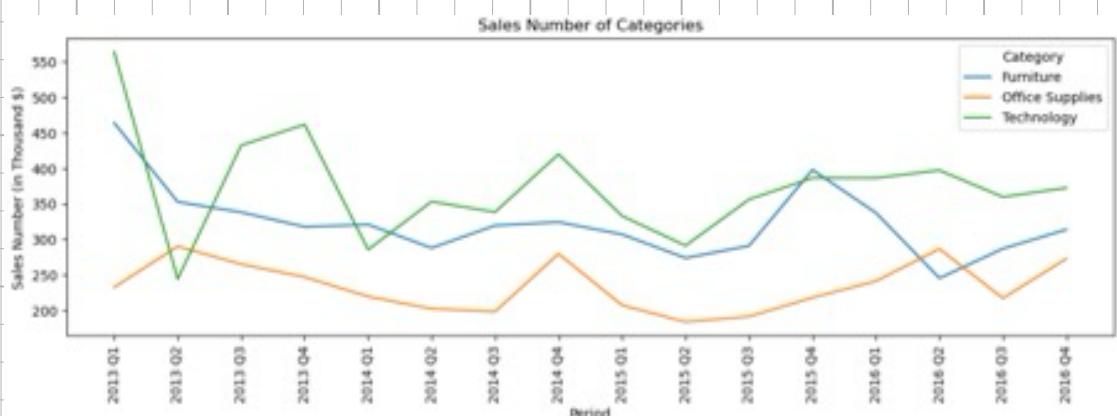
Requirement: Post Data lên Kaggle (public dataset)



Mô tả đặc điểm của dữ liệu cũ dùng nhóm

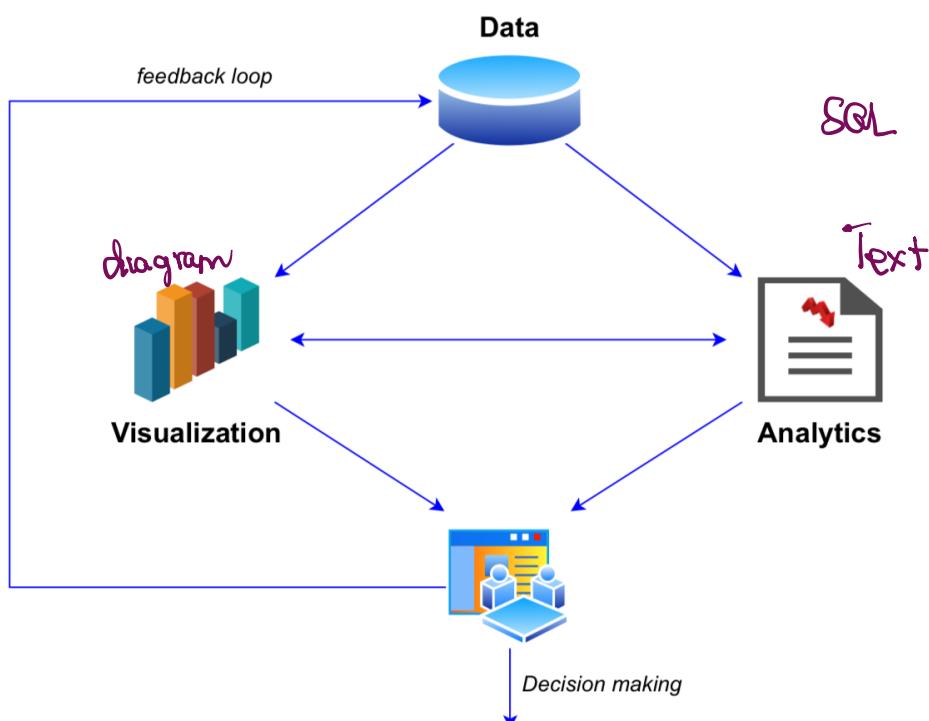
Consider the following table, which shows sales numbers for three categories, by quarter, over a four-year period. What trends can you see?

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	\$463,988	\$352,779	\$338,169	\$317,735	\$320,875	\$287,934	\$319,537	\$324,319
Office Supplies	\$232,558	\$290,055	\$265,083	\$246,946	\$219,514	\$202,412	\$198,268	\$279,679
Technology	\$563,866	\$244,045	\$432,299	\$461,616	\$285,527	\$353,237	\$338,360	\$420,018
Category	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4
Furniture	\$307,028	\$273,836	\$290,886	\$397,912	\$337,299	\$245,445	\$286,972	\$313,878
Office Supplies	\$207,363	\$183,631	\$191,405	\$217,950	\$241,281	\$286,548	\$217,198	\$272,870
Technology	\$333,002	\$291,116	\$356,243	\$386,445	\$386,387	\$397,201	\$359,656	\$375,229



Visualization systems (vis) provide visual representations of datasets designed to help people carry out tasks more effectively.

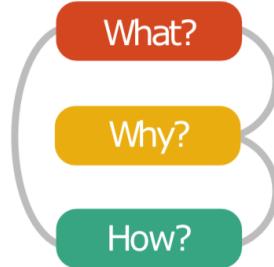
- Computer-based visualization systems



Why analyze?

Three-part analysis framework for a vis instance:

- **what** data is shown in the views
- **why** is the task being performed
- **how** is the vis idiom constructed in terms of design choices.



Ordered data can be

- **sequential**, where there is a homogeneous range from a minimum to a maximum value,
- **diverging**, which can be deconstructed into two sequences pointing in opposite directions that meet at a common zero point
- **cyclic**, where the values wrap around back to a starting point rather than continuing to increase indefinitely.

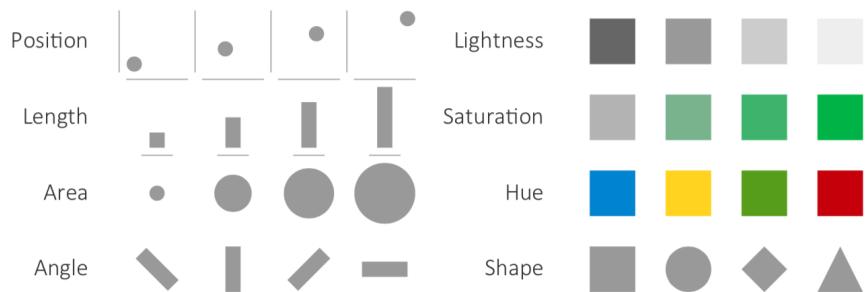
- Describe the nature of information within the values assigned to variables

Type	Measure property	Mathematical operators	Advanced operations	Central tendency	Variability
Nominal	Classification, membership	=, ≠	Grouping	Mode	Qualitative variation
Ordinal	Comparison, level	>, <	Sorting	Median	Range, interquartile range
Interval	Difference, affinity	+, −	Comparison to a standard	Arithmetic mean	Deviation
Ratio	Magnitude, amount	*, /	Ratio	Geometric mean, harmonic mean	Coefficient of variation, studentized range

Visual Encoding

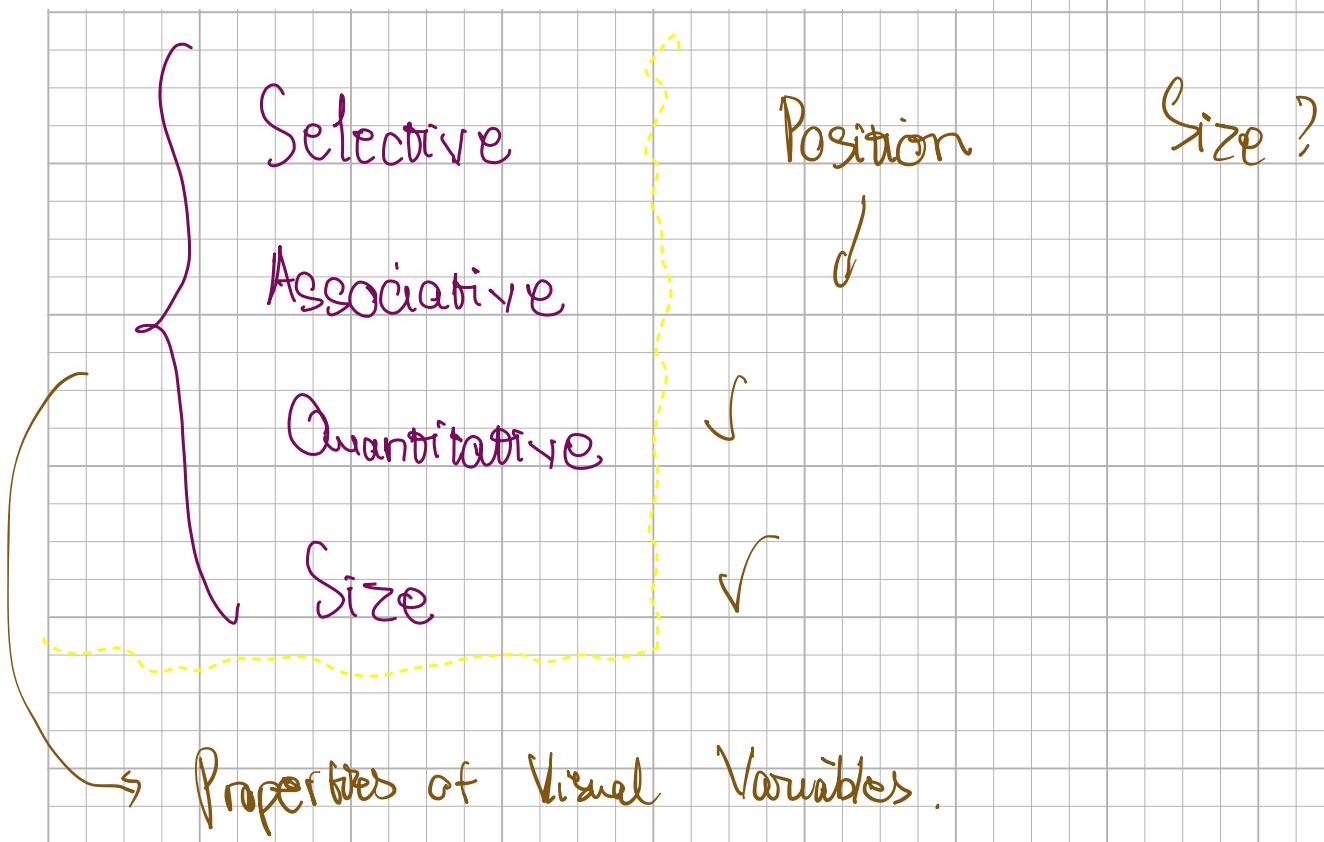
Concept 2

Visual variables (aesthetics, channels) such as position, shape, or hue, which control appearance of marks.



VISUAL VARIABLES

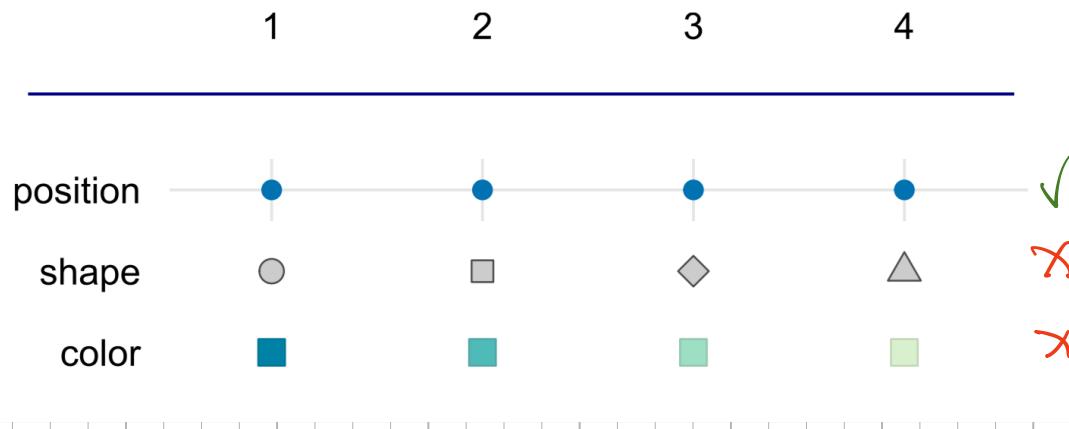
Variables	Changes
Position	changes in the x, y, location
Size	changes in length, area
Shape	infinite number of shapes
Value (lightness, saturation)	changes from light to dark
Hue	changes in hue at a given color
Orientation (angle)	changes in alignment
Texture	variation in 'grain'



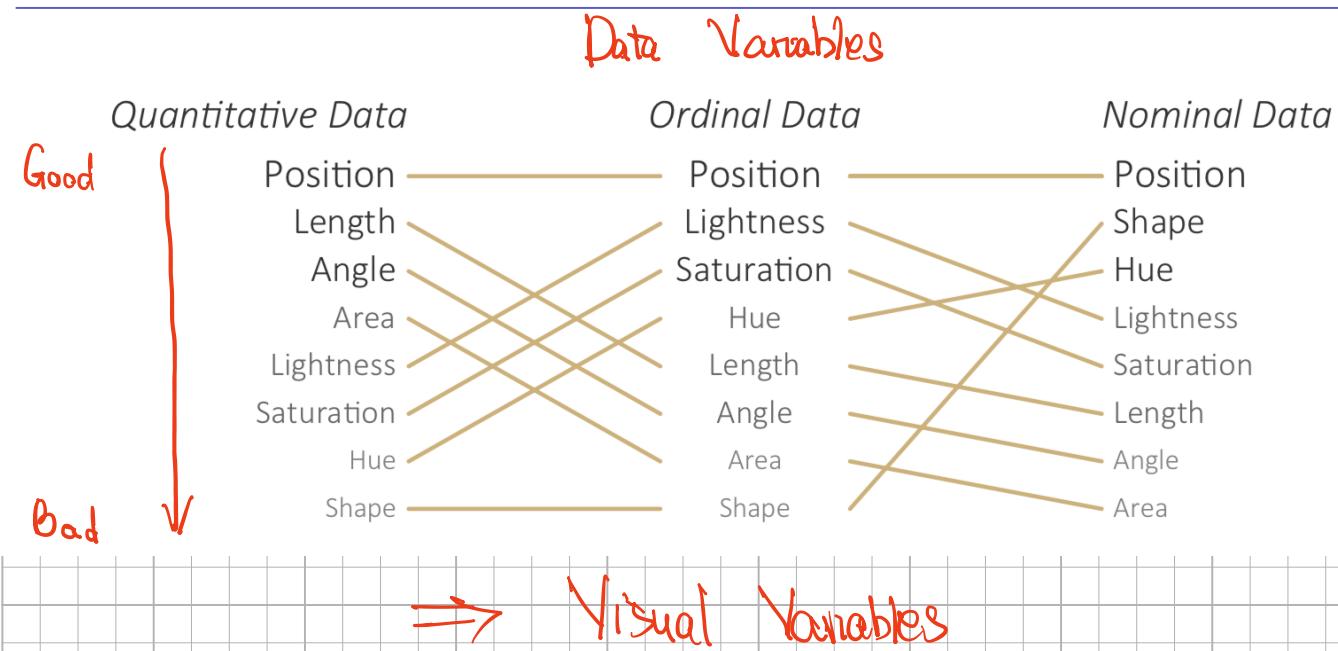
FROM DATA VARS TO VISUAL VARS

Mapping Data Variables onto Visual Variables (cont.)

- To map data values onto aesthetics, we need to specify which data values correspond to which specific aesthetics values.
- This mapping between data values and aesthetics values is created via **scales**, a scale must be one-to-one.

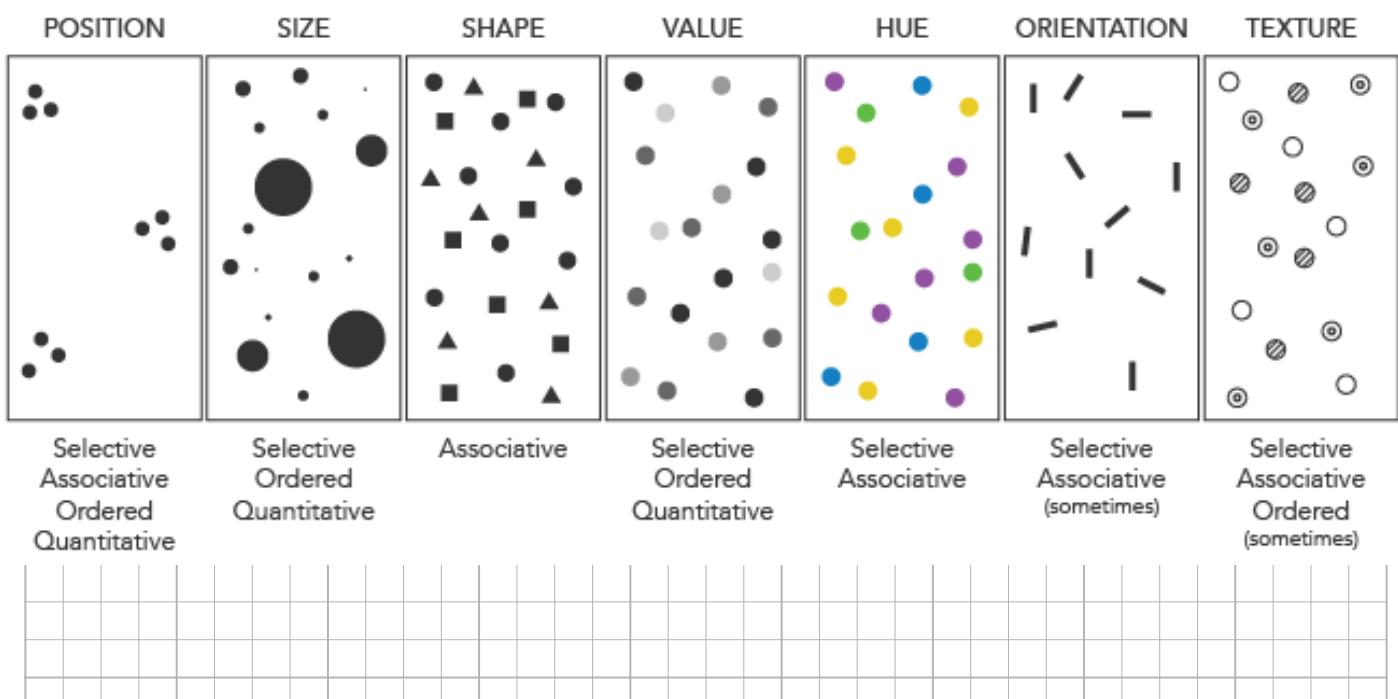


Effectiveness Ranking Of Visual Variables

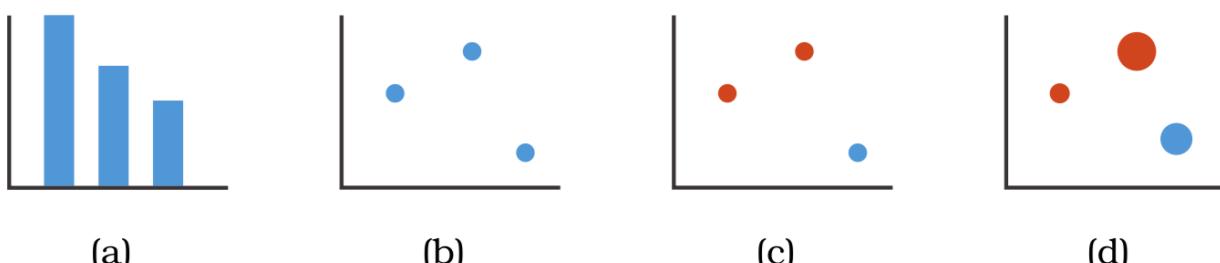


Color is with high associative property.

Bertin's Visual Variables



- (a) **Bar chart** encode two attributes using a line mark with the vertical spatial position channel for the quantitative attribute, and the horizontal spatial position channel for the categorical attribute.
- (b) **Scatterplot** encode two quantitative attributes using point marks and both vertical and horizontal spatial position.
- (c) A third categorical attribute is encoded by adding color to the scatterplot.
- (d) Adding the visual channel of size encodes a fourth quantitative attribute as well.

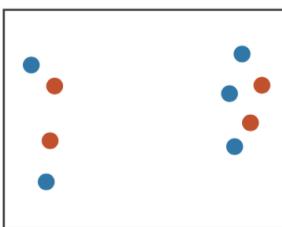


Separability vs. Integrality



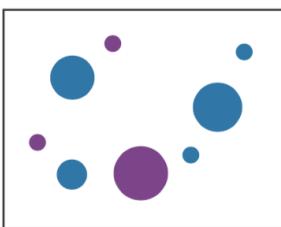
- We cannot treat all visual channels as **completely independent** from each other, because some have **dependencies** and **interactions** with others

Position
+ Hue (Color)



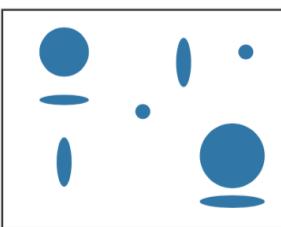
Fully separable

Size
+ Hue (Color)



Some interference

Width
+ Height



Some/significant interference

Red
+ Green



Major interference

Popout

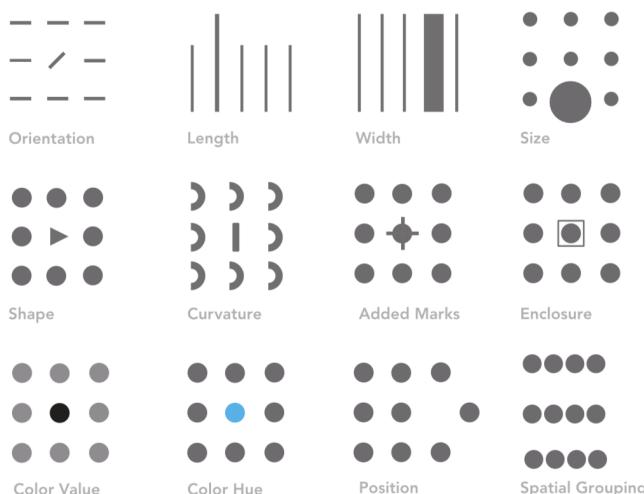


Many visual channels provide visual **popout**, where a distinct item stands out from many others immediately

- Task: Find the **red** dot
- Question: how long does it take?
- *Parallel processing* on many individual channels
 - speed independent of distractor count
 - speed depends on channel and amount of difference from distractors
- *Serial search* for (almost all) combinations
 - speed depends on number of distractors

Popout (cont.)

- Many channels support visual popout



The most effective

pop-out channel

is Color

Seminar:



Mathematics For Machine Learning

Visual perception

1) proximity →

2) closure

3) Similartiy

4) Continuity . Continuity

5) Enclosure

6) Connection

Design

I) Braille

II) Dashboard

III) Graffiti

IV - Table

VISUAL PERCEPTION

① Những thành phần giống nhau thì đặt gần nhau.

* * . * * * .
* * * 4 * * * 0
* * * 0 * * * 0

⇒ Theo dòng

* * * * *
* * * * *
* * * * *
* * * * *
* * * * *

⇒ Theo Col.

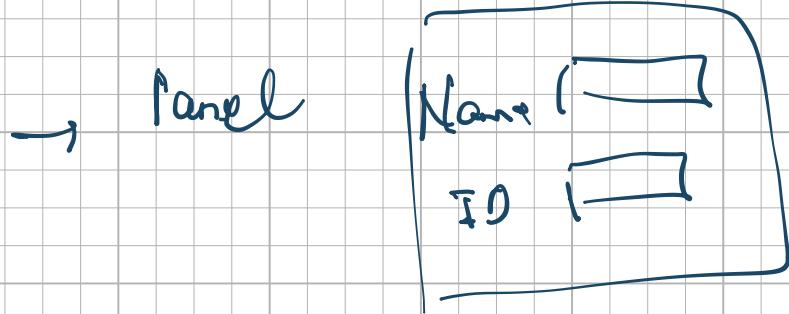
③ Thuộc tính.

- Mô hình \Rightarrow Encode thông tin.
- Hình ảnh \Rightarrow Chất lượng of dashboard
- Koch music

⇒ Sử dụng chung 1 cách encode giữa các biến đó.

5) Enclosure

- Alarms
- Tao chuy

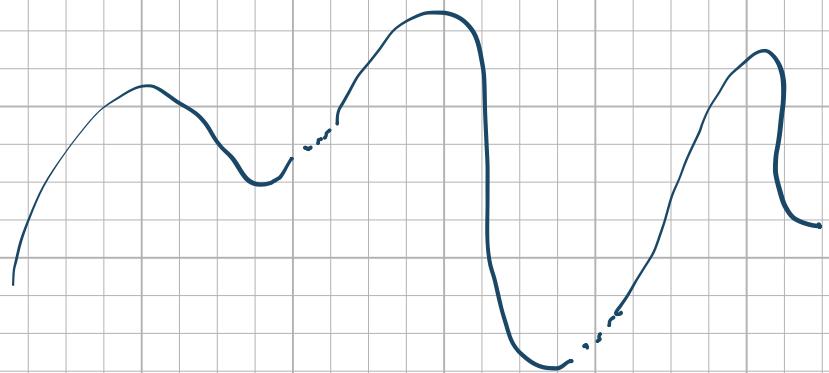


2) Closure

Tôi suy diễn , Né (CN) material design :

→ How gran?

4) Continuity



Line chart

Dot Chart

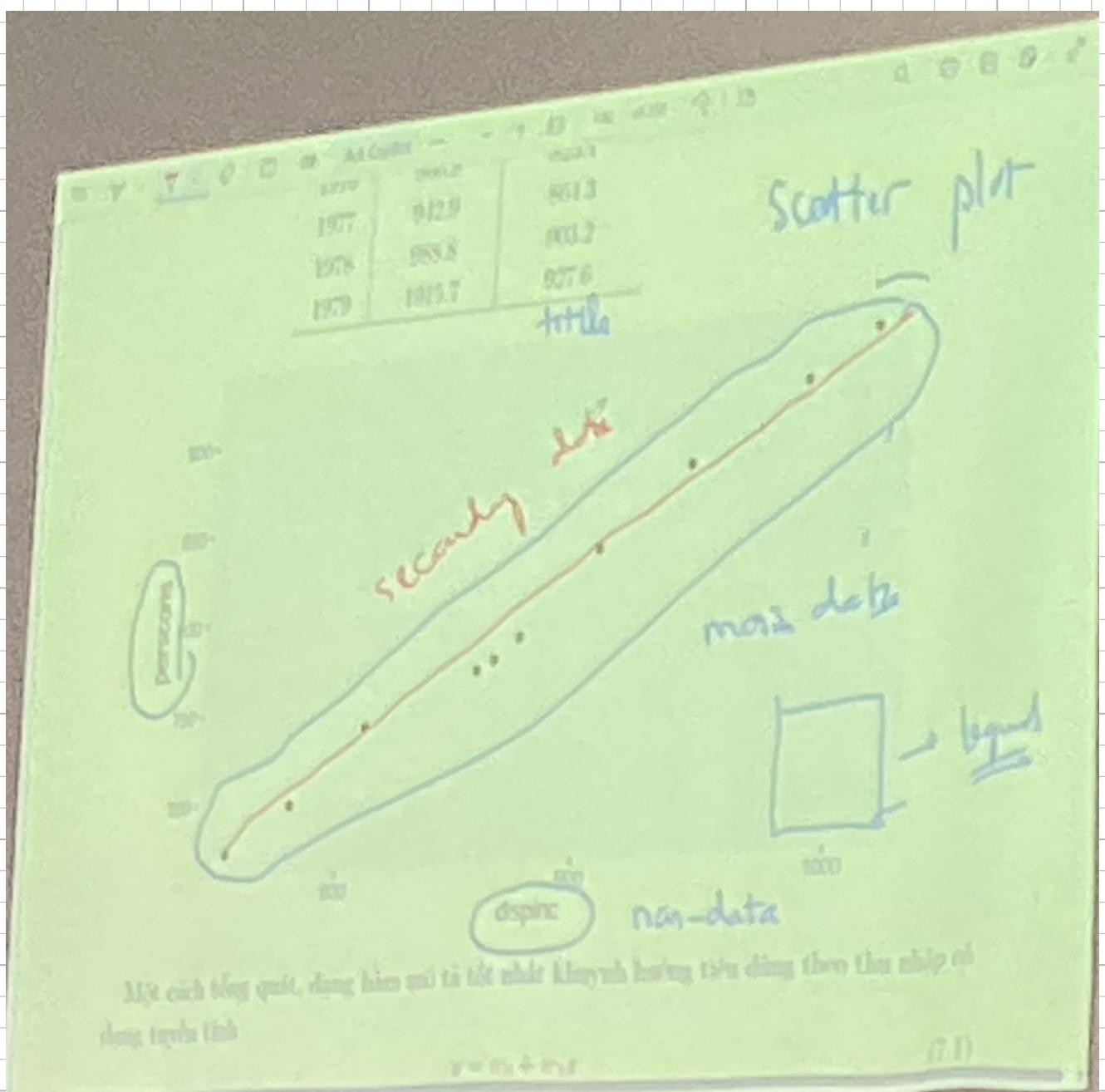
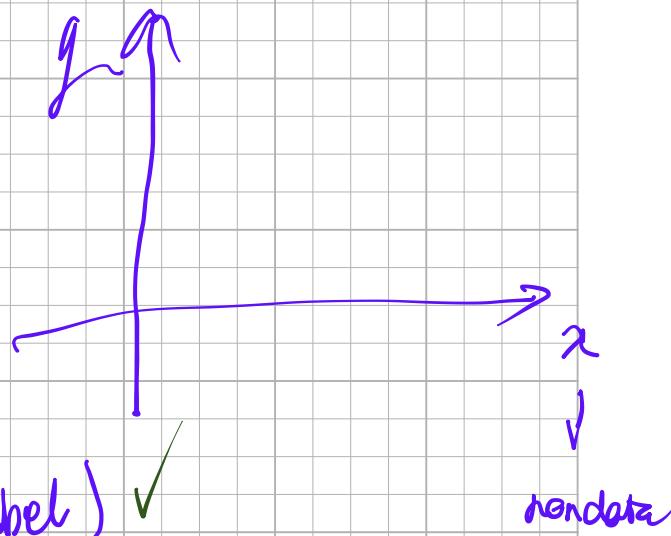
5) Connection.

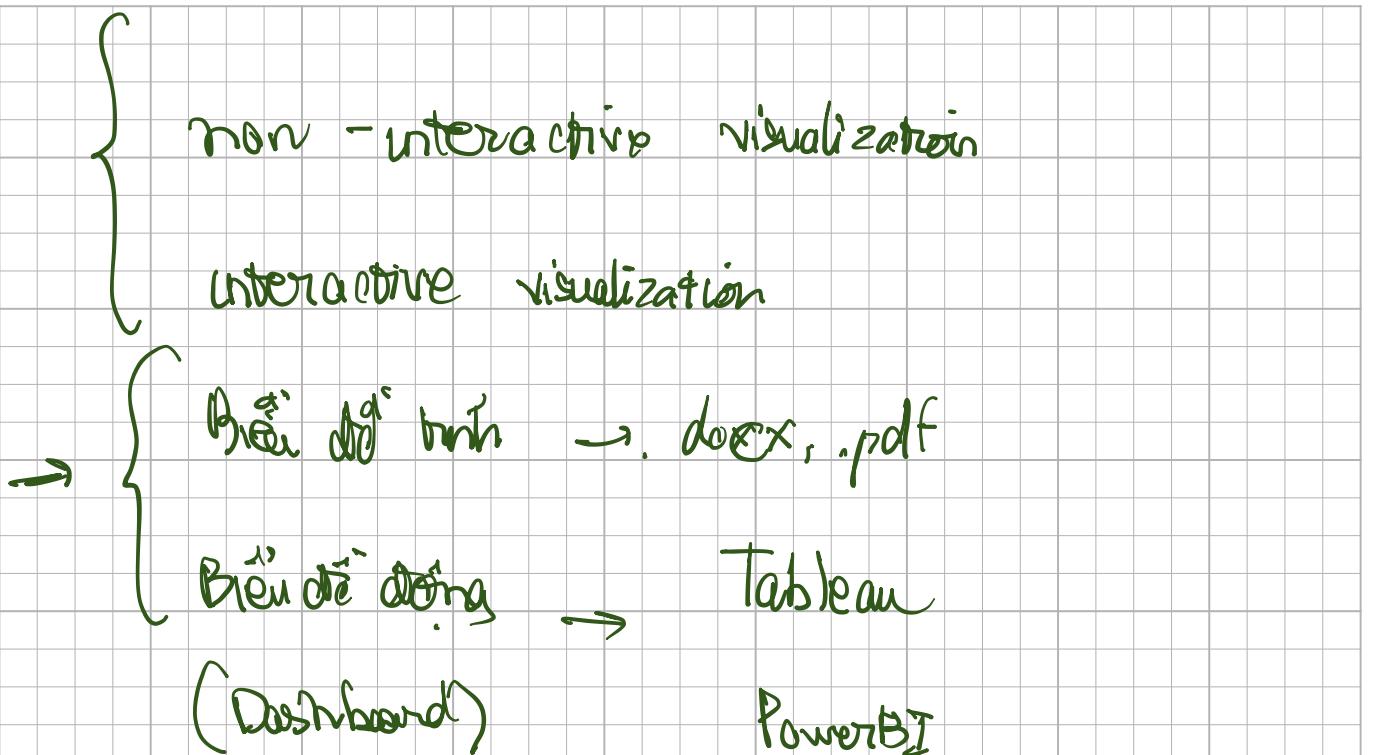
In network graph

DESIGN

I - Brief dc:

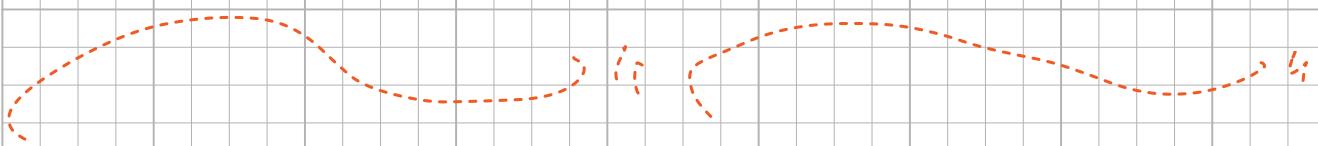
{ Main data ✓
Secondary data
Non-data (label) ✓





Chất lượng biểu đồ đánh giá theo các tiêu chí:

- Khoa: Bitmap (.png, .jpg)
- Tốt: Vector (.svg) (Latex → .pdf)

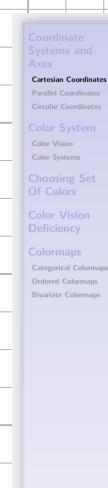


COORDINATE SYSTEM AND AXES

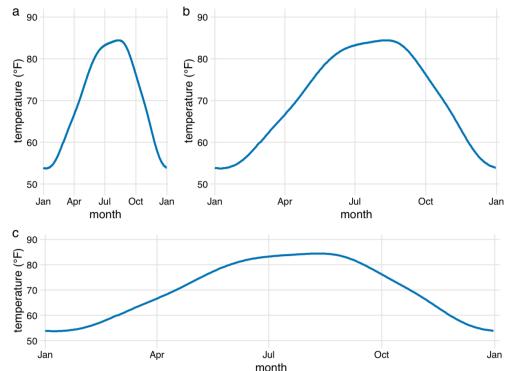
Cartesian

Parallel

Circular



Units (cont.)



6

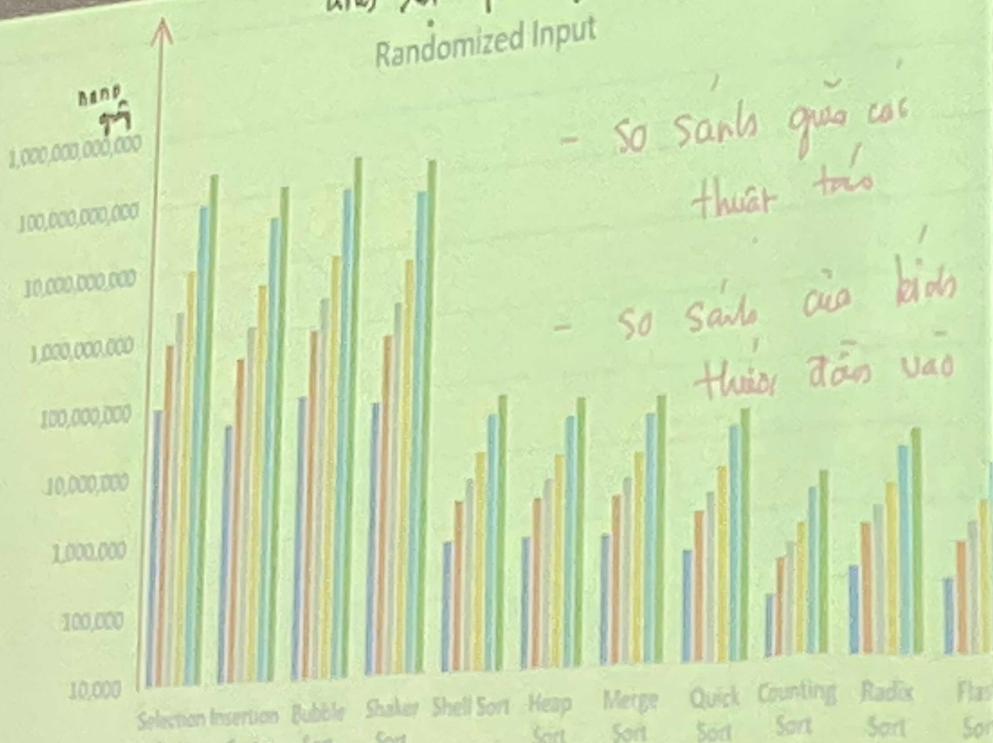
Emphasize axis

Ảnh xà phi tuyến.

anh xà phi tuyến
Randomized Input

- So Sánh giữa các
thuật toán

- So Sánh của kích
thướt, độ sâu vào



Nonlinear Axes



- The most commonly used nonlinear scale is the **logarithmic scale**

