

Data Object & Attribute Types

ตัวอย่างการบันทึกข้อมูล

- Relational records
- Relational tables, highly structured
- Data matrix, e.g., numerical matrix, crosstabs

	China	England	France	Japan	USA	Total
Active Outdoors Crochet Glove	12.00	4.00	1.00	240.00	257.00	
Active Outdoors Lycra Glove		10.00	6.00		323.00	339.00
Influx Crochet Glove	3.00	6.00	8.00		132.00	145.00
Influx Lycra Glove		2.00			143.00	145.00
Triumph Pro Helmet	3.00	1.00	7.00		233.00	344.00
Triumph Verigo Helmet		3.00	22.00		474.00	499.00
Xtreme Adult Helmet	8.00	8.00	7.00	2.00	251.00	276.00
Xtreme Youth Helmet		1.00			76.00	77.00
Total	14.00	43.00	54.00	3.00	1,972.00	2,086.00

Record DATA

	First Name	Last Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

no relation

Car_ID	Model	Year	Value	Pers_ID
101	Bentley	1973	100000	0
102	Rolls Royce	1965	330000	0
103	Peugeot	1993	500	3
104	Ferrari	2005	150000	4
105	Renault	1998	2000	3
106	Renault	2001	7000	3
107	Smart	1999	2000	2

- Transaction data

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

	season	timout	lost	win	game	score	ball	ply	coach	team
Document 1	2	0	2	0	6	2	0	5	0	3
Document 2	0	0	3	0	0	1	2	0	7	0
Document 3	0	3	0	2	2	1	0	0	1	0

Spatial, image & multi

Spatial data: maps

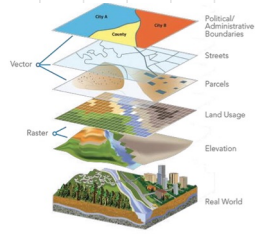


Image data: *เชิงพื้นที่* *เชิงเวลา*

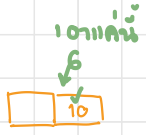
Video data: *Spatio-temporal*

Data Objects

- Data sets are made up of data objects
- A data object represents an entity
- Examples:
 - sales database: customers, store items, sales
 - medical database: patients, treatments
 - university database: students, professors, courses
- Also called *samples*, *examples*, *instances*, *data points*, *objects*, *tuples*
- Data objects are described by **attributes**
- Database rows → data objects; columns → attributes

ลักษณะที่สำคัญของ DATA

- Dimensionality
 - Curse of dimensionality
- Sparsity *สนใจแต่ตรงที่มีข้อมูล*
 - Only presence counts
- Resolution
 - Patterns depend on the scale
- Distribution
 - Centrality and dispersion



Attributes

- Attribute (or dimensions, features, variables)
 - A data field, representing a characteristic or feature of a data object.
 - E.g., *customer_ID*, *name*, *address*
- Types:
 - Nominal (e.g., red, blue)
 - Binary (e.g., {true, false})
 - Ordinal (e.g., {freshman, sophomore, junior, senior})
 - Numeric: quantitative
 - Interval-scaled: 100°C is interval scales
 - Ratio-scaled: 100°K is ratio scaled since it is twice as high as 50°K
- Q1: Is student ID a nominal, ordinal, or interval-scaled data?
- Q2: What about eye color? Or color in the color spectrum of physics?

*** ตัวอย่างอุปมาอุปไมยในทางปฏิบัติ ***

การจัดแบ่งประเภทของข้อมูลในลักษณะ

ก. ตามลักษณะของข้อมูล

- สเกลนามบัญญัติ (Nominal Scale)
- สเกลเรียงลำดับ (Ordinal or Ranking Scale)
- สเกลอันดับภาค (Interval Scale)
- สเกลอัตราส่วน (Ratio Scale)

ข. ตามการจัดเก็บข้อมูล

- แสงปฐมภูมิ *ขึ้นกับกำหนดโดยเรา*
- แสงทุติยภูมิ *ขึ้นกับจากผู้อื่นรวมแล้ว, มีตัวต้น*

ค. ตามช่วงเวลาอ้างอิงของข้อมูล

- อนุกรมเวลา
- ภาคตัดขวาง

Cross-sectional Data *ข้อมูล ณ เวลาใดเวลาหนึ่ง*

Each observation represents a different individual unit (e.g., person) at the same point in time (e.g., monthly VISA balances).

We are interested in

- variation among observations or in relationships.

We can combine the two data types to get pooled cross-sectional and time series data.

Time Series Data

- Each observation in the sample represents a different equally spaced point in time (e.g., years, months, days).
- Periodicity may be annual, quarterly, monthly, weekly, daily, hourly, etc.
- We are interested in trends and patterns over time (e.g., annual growth in consumer debit card use from 2001 to 2008).

24 Wichittha Chaisriwongthot

4) ระดับข้อมูลของการวัด (Data Levels of Measurement)

เชิงคุณภาพ (Qualitative) *คุณลักษณะ, ไม่วัด*

1. สเกลนามบัญญัติ (nominal scale)
2. สเกลเรียงลำดับ (ordinal scale)

เชิงปริมาณ (Quantitative) *วัดได้, วัดค่า*

3. สเกลอันดับภาค (interval scale) → 0 ไม่เท่า 100 องศาเซลเซียส
4. สเกลอัตราส่วน (ratio scale) → 0 เท่า

คุณสมบัติ	สเกลนามบัญญัติ	สเกลเรียงลำดับ	สเกลอันดับภาค	สเกลอัตราส่วน
1. จำแนกความแตกต่าง	✓	✓	✓	✓
2. เรียงลำดับ	✗	✓	✓	✓
3. การ +, -	✗	✗	✓	✓
4. การ ×, ÷	✗	✗	✗	✓
5. มีศูนย์แท้	✗	✗	✗	✓

Operations allowed	Categories only or Only Counting (เช่น ความถี่, Mode)	Rank has meaning or Counting and order (เช่น Mode Median, rank test)	Distance has meaning and Use sums of differences (เช่น Mean, SD)	Meaningful zero exists All Statistical Operations, ratios of number
--------------------	---	--	--	---

23 Wichittha Chaisriwongthot

Discrete vs. Continuous Attributes

- Discrete Attribute
 - Has only a finite or countably infinite set of values
 - E.g., zip codes, profession, or the set of words in a collection of documents
 - Sometimes, represented as integer variables
 - Note: Binary attributes are a special case of discrete attributes
- Continuous Attribute
 - Has real numbers as attribute values
 - E.g., temperature, height, or weight
 - Practically, real values can only be measured and represented using a finite number of digits
 - Continuous attributes are typically represented as floating-point variables

Basic Statistical Descriptions of Data
ใช้คำกลางทางสถิติ → เป็นคำกลางของข้อมูล [Mean]