

# chapter 2. Getting to know your data.

## Data

1D

1	1	12	2	5
2	2	11	7	2
1	1	15	9	3
0	0	10	1	-3
-1	-1	20	12	-2
1	1	19	6	-5

2D

	1	1	12	2	5
1	2	2	11	7	2
2	1	1	15	9	3
1	0	0	10	1	-3
1	0	-1	20	12	-2
0	-1	1	19	6	-5
-1	1	1	19	6	-5
1	1	1	19	6	-5

	1	1		12		2		5	
1		2		2		11		7	2
2	1		1		15		9		3
2	1		0		10		1		-3
1	0		-1		20		12		-2
0	-1		1		19		6		-5
-1	1			19		6			-5
1				19		6			-5

	1	1		12		2		5
	1	2	2		11		7	2
1	2	1	1		15		9	3
2	1	0	0		10		1	-3
1	0	-1			20		12	-2
0	-1	1			19		6	-5
-1	1			19		6		-5
1								1

	1	1		12		2		5
1		2	2		11		7	2
	2	1	1		15		9	3
2		1	0		10		1	-3
1	0		-1		20		12	-2
0		-1	1		19		6	-5
-1		1		19		6		-5
1								

10 จอสี แยกช่วง or ภา

4D

ถ้าเป็น 5D... บลาๆๆ  
เรียงซ้อนกันไปเรื่อยๆ

## Types of Data Sets: (1) Record Data

หลายตารางที่สัมพันธ์กัน

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

	China	England	France	Japan	USA	Total
Active Outdoor Crochet Glove	12.00	4.00	1.00	240.00	257.00	
Active Outdoor Lycra Glove		10.00	6.00		323.00	339.00
Infelix Crochet Glove	3.00	6.00	8.00		132.00	149.00
Infelix Lycra Glove		2.00			143.00	145.00
Triumph Pro Helmet	3.00	1.00	7.00		333.00	344.00
Triumph Vertigo 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

Person:

Pers_ID	Surname	First_Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

Car:

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

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

ทำให้อ่านเข้าใจ  
ตารางซ้อนกัน 4x

- Document data: Term-frequency vector (matrix) of text documents

## Types of Data Sets: (4) Spatial, image and multimed

- 

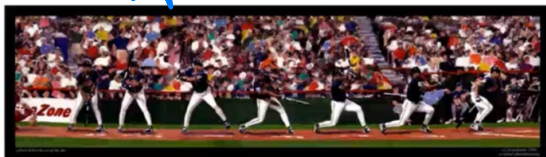
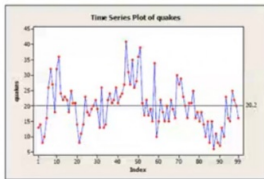


การ  
ใช้ประโยชน์

- Molecular Structures *1500*
- Social or information networks

Video data: **spatio-temporal** → ข้อมูลที่มีเวลา  
ข้อมูลเชิงวิถิ์  
มาเก็บพร้อม

- Video data: sequence of images
- Temporal data: time-series



ข้อมูลที่มีเวลาไม่กี่ชั่วโมง

- Sequential Data: transaction sequences  
ลำดับข้อมูลไปเรื่อยๆ  
Ex. DNA
- Genetic sequence data

Ex. DNA



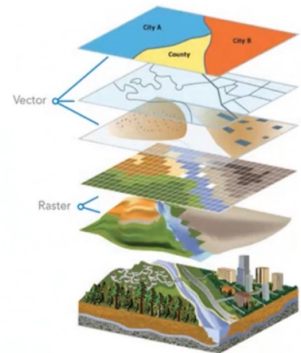
- ❑ Spatial data: maps



- Image data:

- ❑ Video data: Spatio-temporal

## ข้อมูลเชิงวิสัย



## Data Objects

- ❑ Data sets are made up of data objects → หลายๆอันมารวมกัน
- ❑ A **data object** represents an entity ข้อมูลแต่ละตัว
- ❑ Examples:

ข้อมูล แหล่งที่มา

- 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

attributes

คุณสมบัติที่อธิบายข้อมูลแต่ละตัว

# Attributes

ชื่อเรียก

- ❑ **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}) มีแค่ 2 ค่า true, false ✓
  - ❑ Ordinal (e.g., {freshman, sophomore, junior, senior}) → ข้อมูลที่เรียงลำดับ
  - ❑ Numeric: quantitative → ข้อมูลตัวเลข → +, -, ×, ÷ แล้วมีความหมาย
    - ❑ 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?